



Perspectives On...

• On the Subject of Subjects

by Arlene G. Taylor

• Library Benchmarking: Old Wine in New Bottles?

Sarah M. Pritchard

On the Subject of Subjects

by Arlene G. Taylor

This was a period of much argument over the relative virtues of classified and dictionary catalogs not only among librarians but among readers and scholars in general and even in reports to the House of Commons. Feelings ran very high on the subject, and rather emotional arguments came forth on both sides of the issue, from the statement that classified catalogs and indexes were not needed because living librarians were better than subject catalogs to the opinion that any intelligent man who was sufficiently interested in a subject to want to consult material on it could just as well use author entries as subject, for he would, of course, know the names of all the authors who had written in his field.¹

—Ruth French Strout

The time and place Strout refers to is in the 1830s in Great Britain. It seems that subject cataloging has a long history of being questioned. Even though Charles Cutter convinced American librarians to use subject headings in dictionary catalogs by the beginning of this century, the concept of subject searching remained a disreputable one. Many catalog use studies showed that most searches were for known items, or at least were for a known author. Even though some studies showed a majority of subject searches, especially in public libraries, these tended to be ignored.² Thus, in the 1980s when the Council on Library Resources sponsored a national study of online catalogs, it came as something of a shock to people that a majority of searching done in online catalogs was subject searching. Even so, subject cataloging continues to be disparaged by many librarians.

Brad Young tells of getting on an elevator at an American Library Association (ALA) Conference a few years ago, carrying a copy of the article entitled "Can Subject Headings Be Saved?" Another librarian looked at the title and said, "Who cares?"³ A similar attitude was exhibited by reference librarians in response to a listserv message written by Thomas Mann, a reference librarian at the Library of Congress. In his message Mann

suggested that librarians teaching bibliographic instruction should point out the helpfulness of using narrower term references in the "red books" (i.e., *Library of Congress Subject Headings*); the usefulness of narrower terms that are alphabetically adjacent to broader terms in the "red books;" the possibility of finding appropriate subject headings in the tracings on a relevant record found in the catalog; and the possibility of narrowing a search in a catalog by examining the array of subdivisions given after a more general subject heading. Respondents indicated their disdain for Library of Congress Subject Headings, commenting that they recommend keyword searching instead.⁴ I will say more about keyword versus controlled vocabulary later.

Soon after the national online catalog study, Ray Larson began an extensive study tracking the use of the MELVYL online catalog system over a period of six years.⁵ He found that subject index use, as a percentage of index usages, declined over the six-year period. Several librarians latched on to this finding with seeming glee. For example, Peter Graham wrote:

I have placed subject access low on this list of priorities. In spite of OPAC-use study claims of the early 1980s (e.g., Joseph Matthews, *Public Access to Online Catalogs*, 2nd ed. New York: Neal-Schuman, 1985), I have considerable doubt that librarians can effectively provide good enough subject access to warrant the present great costs, nor do I think patrons are aware of costs when they are asked if they like the idea. Recent more tempered comments on the need for subject access have appeared, e.g., Ray N. Larson's presentation at an ALA 1988 Summer Conference meeting of the Library Research Round Table, "The Decline of Subject Searching in Online Public Access Catalogs"....⁶

However, as Larson pointed out in his write-up of the research⁷ and also noted at an ALA Conference presentation in 1994,⁸ it was obvious where the decline in subject index use percentages was being made up: in the title keyword index. That is, users were still trying to do subject searching, but, because they knew so little about the controlled vocabulary, they did not know how to search it. Just under half of subject index searches produced zero results (i.e., users could not match their vocabulary with that of the system), and most of the remaining searches produced too many retrievals (i.e., search terms used were so broad that 200 or more items were retrieved—such an overwhelming number, in many cases, that the user did not even start browsing through them). Users try to correct for these failures with title keyword searching. As Larson concludes, "The subject index,

Arlene G. Taylor, School of Library & Information Science, University of Pittsburgh, 135 N. Bellefield Ave., Pittsburgh, Pennsylvania 15260

even after the decline discussed above, is still one of the most commonly used search access points in the online catalog."⁹

SUBJECT ACCESS AND THE INTERNET

It seems fairly clear at this point that people are doing a lot of subject searching, especially on the Internet; however, there is considerable evidence that subject searching of the Internet is not as successful as many would wish. The Internet has been likened to a number of different metaphors: e.g., a mansion with many rooms where the lights have not been turned on yet; or, an enormous bookstore with books stacked on tables, chairs, and floor, and new ones being continuously added helter-skelter to the piles. Michael Gorman has added to the bookstore example that all the volumes have had their front matter torn off so that, in addition to not having any order, one cannot assess the validity of who supposedly wrote them or whether they were ever given a title.¹⁰ Robert Jacobson has asked us to imagine a library where the books have all been donated by patrons and placed randomly on the shelves and where new collections arrive daily; patrons can roam the stacks freely and move the volumes around from shelf to shelf whenever they want. He goes on to say that, yes, the network is supported by browsers, searching tools, and informal indexes, many of which are linked and reachable with the click of a mouse. But then, paraphrasing David S. Magier, a Columbia University librarian, he points out that most of the tools and guides were created by computer experts, not librarians, and as a result researchers can cruise cyberspace like anyone else but typically cannot zero in on specific, high-quality resources. What is needed, he says, is what subject librarians have always provided for non-electronic resources.¹¹

Stuart Whitwell begins a recent article with "Communities are often best defined by the threat from outside."¹² James Beach, after quoting Whitwell, suggests that perhaps information professionals finally have a common threat around which we can unite. That threat, he suggests, is chaos.¹³ "Chaos" would seem to be a fairly good description for the net; but librarians are beginning to rally to the challenge. Francis Miksa says that what librarians have developed in the way of controlled vocabulary and classification is a "sense-making operation."¹⁴ And now librarians are beginning to put some sense-making into the Internet. One librarian, Karen G. Schneider, challenges others this way: "The fact that these hideous hotlists are all the rage is proof positive that we need more librarians active on the net! If you're a librarian, and you're treating the Internet like a database to dip into rather than a community to participate in, I exhort you to get off your duff and PARTICIPATE!"¹⁵ Schneider characterizes subject guides on the net this way:

I recommend starting a search with a subject guide because keyword search engines can't duplicate the ability of skilled humans to evaluate and organize information. Which subject guides? Start with the librarian-built guides, which for the most part are selective and well-annotated. You will save time and benefit from the judgment of these experienced hunters and gatherers.¹⁶

Some of the librarian-built guides are efforts of individuals, while others are collaborative, sometimes institutional, efforts. For example, persons at four research institutions in the New York City area (at the instigation of David S. Magier), are exploring, categorizing, and evaluating Internet resources in eight fields: area studies, art and architecture, business, history, literature, music and performing arts, science, and social sci-

ence. The idea is to divide up the work into manageable pieces and not duplicate each other.¹⁷

Some research on organization of the Internet is being done. C. Olivia Frost and Joseph Janes have undertaken research comparing users' success with three different Gopher menu structures: one organized by broad subject area, one organized by Dewey Decimal Classification, and one organized by Library of Congress Classification.¹⁸ Although data are not yet completely analyzed, findings from the study will give insight into the ways people search in the network environment.

Recently, the National Science Foundation gave major funding to six projects working on creating "digital libraries," and the title of one news article about this is quite telling: "Turning an Info-Glut into a Library."¹⁹ The projects are all quite different from each other, working with text, images, or both; but the ultimate goal will be to create giant multimedia libraries. And all incorporate some type of subject organization. Another project is one begun by OCLC a few months ago to catalog Internet resources. In this project catalogers in cooperating libraries are cataloging documents found on the Internet that seem to be appropriate additions to that library's catalog. The resulting catalog records are standard MARC records complete with the controlled vocabulary required by the cataloger's institution and containing a field for the URL (Uniform Resource Locator) of the document being cataloged. These records are being entered into the OCLC database and will become part of NetFirst, "a comprehensive database of Internet-accessible resources that will be a practical, easy-to-use tool for Internet exploration and discovery."²⁰ The announcement for NetFirst makes a point of saying that the database citations will contain summary descriptions and subject headings.

Subject Access and Standards

Information specialists are also creating standards for future cataloging, and these standards include places for subject data. The Library of Congress, with the cooperation of OCLC and the Research Libraries Group, has taken the lead in finding a way to increase the number of mutually acceptable bibliographic records available for use. One outcome of the cooperative project begun in 1992 has been a set of "Core Records." Records done to this standard are less than full MARC records but provide a basic minimum in which a later user can have confidence that certain elements are present, and that prescribed ones are authority controlled.²¹ The parts particularly relevant here concern classification and subject headings. The Core Record for monographs requires at least one classification number from an established classification system recognized by USMARC and, if appropriate, at least one or two subject headings at the appropriate level of specificity from an established thesaurus or subject heading system recognized by USMARC. Core Records also have been defined for musical scores, sound recordings, non-roman script monographs, and serials. A task group is working on one for audiovisual materials.

Another standard for access to documents is the TEI Header. The Text Encoding Initiative (TEI) is an international project that has developed a standard for the SGML mark-up of electronic texts for scholarly research. Under the TEI Guidelines, every electronic text that is encoded according to the guidelines has at its beginning the TEI Header, which is the electronic equivalent to the front matter of a printed book combined with the equivalent of an electronic codebook. The information required for a TEI Header resembles a MARC record in many respects. It is divided into four parts: File Description, Encoding

Description, Profile Description, and Revision Description. Within the Profile Description are positions for "<keywords>" and "<classCode>." Each of these can be identified as coming from a standard thesaurus or classification, but this is not required. The intent is that encoders provide as much of the header information as possible, even in prose if necessary, and that it be left to catalogers, archivists, and others to create the structured subelements.²²

Seeing a need for a finding mechanism in between the extremes of the minimal information provided by automatically generated indexes such as Lycos and WebCrawler and the considerable detail of MARC Records and the TEI Header, OCLC and the National Center for Supercomputing Applications convened a Metadata Workshop, in March 1995, to try to define a metadata record to mediate the two extremes and be able to be enhanced or mapped to more complex records as needed. The result is being called the Dublin Core (the meeting having been held at OCLC in Dublin, Ohio). The hope is that the 13 elements defined are simple enough that any creator of an electronic document would be able to provide the record. All the elements are optional, but the first one listed is "Subject," giving a sense that the participants in the Metadata Workshop consider subject to be of prime importance. The report from the Workshop defines subject: "The Subject is the field of knowledge to which the work belongs. This may be a general description of a broad discipline, or a series of descriptors of differing scope." The report goes on to specify how to qualify the subject element by a scheme and includes examples of both thesauri/subject heading lists and classification schemes. If no scheme is specified, the element may contain a free text phrase that describes the intellectual content.²³

CONTROLLED VOCABULARY VS. KEYWORDS

Both the TEI Header and the Dublin Core allow subject terminology that is not controlled vocabulary, and automatically generated indexes consist solely of uncontrolled vocabulary. Is this sufficient and satisfactory? Many writers think not, although it depends somewhat on one's reasons for doing a search. If one just wants to find *something* regardless of source or quality, keyword will probably do it. However, if one is a researcher looking for *the best* on a subject or *everything* on a subject, keyword is quite chancy.

It has been pointed out by many writers in many contexts that keyword searching gives the user no assistance in identifying synonyms or near synonyms. Thomas Mann gives an example of a concept that has at least nine words/phrases that mean the same thing:

...capital punishment, death penalty, execution, legal execution, penalty of death, pena capitale, peine de mort, smertnaia kazn', and Todesstrafe can all be understood to refer to the same thing. Authors who write on this subject may also use still other words (*A Life for a Life, To Kill and Be Killed, The Ultimate Coercive Sanction*, and so on) to refer to it in the titles of their works. In systems without vocabulary control of synonyms, researchers who think of only one or two words miss most of the material on their subject without realizing they've missed anything.²⁴

A problem in much of the research that has been done with keyword retrieval seems to be that small scientific databases have often been used, and the sciences seem not to have as many ways of saying the same thing as do the social sciences and humanities (as in the Mann example), although F.W. Lancaster uses some

scientific terminology to illustrate the point: e.g., glues, adhesives, cements.²⁵

Another problem with keyword searching is that words taken out of context can have several different meanings. Search engines cannot tell you if a suit is a legal term or a set of clothing. Many early experiments with keyword searching were conducted in databases of documents all from one discipline (or even narrower subject areas) where the ambiguity of terminology was not so apparent.

Finally, keywords give no clues to their relationships with other terms and concepts that are often widely separated alphabetically. It can be suggested to a person looking in a subject list or thesaurus under *lumber* that a related term is *timber* and that one might also think about the broader concept *forests* or the narrower concept *hardwoods*. This is not possible with keyword searching.

Use of controlled vocabulary also has the advantage of offering both the indexer and the user of the same thesaurus similar ideas about what to call certain concepts. Lancaster asserts that:

Indexers are more likely to agree on the terms needed to describe a particular topic if they are selected from a pre-established list than when given a free hand to use any terms they wish. The same is true for the process of searching: It is easier to identify the terms appropriate to some information need if these terms must be selected from a definitive list. The controlled vocabulary, then, tends to bring the language of indexers and of searchers into coincidence.²⁶

This is not to say that there will be indexing consistency in selection of concepts to which indexing terms are to be attached. However, when the same concepts are *identified* in two or more entities, those concepts will have a greater chance of having the same terminology applied to the concepts, regardless of the terminology used by the authors of the documents being indexed.

On the other hand, there are several advantages to free text indexing: low cost; simpler searching; searchability of full information content; no human indexing errors; and no delay in incorporating new terms.²⁷ Researchers and other librarians who think about such things seem to have come to the conclusion that C.P.R. Dubois came to in 1987: "Free text and controlled vocabulary searching can no longer be viewed as antagonistic techniques in information retrieval since they both display advantages and weaknesses dependent on a fairly wide range of context, with the option to use both increasingly favoured."²⁸ Nowhere is this more apparent than on the Internet. As we begin to make some order of the chaos of the Internet, it makes sense that we will let the ephemera be indexed by keyword, and as entities are identified that are obviously of more lasting value, they will be brought into our spheres of controlled vocabulary items so that they can be identified along with the printed materials on the same subject that we will continue to collect for many years.

A final point to consider in relation to controlled vocabulary is that as we incorporate access to pictures and objects, which have no "words" to be searched as "keywords," into the same access systems for books, we attach to these things our words that can be searched. Would it not make sense to use controlled vocabulary for these terms rather than exacerbating the synonym problem by using whatever pops into one's mind at the moment?

Specific Entry

It also makes sense that the items that we do catalog should be done according to a standard, and that standard is likely to

be the Core Record mentioned earlier. The Core Record requires "at least one or two subject headings at the appropriate level of specificity."²⁹ Just as some librarians question the need for controlled vocabulary, some question the need for the practice of "specific entry." This practice stems from Charles Cutter, who wrote:

Enter a work under its subject heading, not under the heading of the class which includes that subject.... Put Lady Cust's book on "the cat" under "Cat," not under Zoology or Mammals, or Domestic animals....³⁰

Later theorists refined Cutter's work. David Haykin wrote:

The heading should be as specific as the topic it is intended to cover. As a corollary, the heading should not be broader than the topic; rather than use a broader heading, the cataloger should use two specific headings which will approximately cover it.³¹

A weakness of the concept of specific entry is that subjects must be described in terms that are constantly changing. If something is cataloged before a suitable term has been added to a standard list, one must choose a term at the next level of generality; so early works on a new topic are sometimes separated from the later ones.

Those who oppose controlled vocabulary usually also oppose specific entry as just another of the idiosyncrasies of anal-retentive catalogers. However, Mann makes an important point from the reference and research point of view:

"Specific entry," as a principle, is particularly important because it makes the choice among many possibly relevant terms much more predictable. That is, when a book could logically be cataloged under several different valid headings at various levels of generality—all of which headings and levels can be found in the LCSH list—specific entry makes the choice of which heading to [search under] more determinable in advance.³²

Specific entry means that if a whole book has been written on, let us say, suicide among people under the age of 20, it should be entered under "Teenagers—Suicidal behavior" and/or "Youth—Suicidal behavior," and should not be entered with the hundreds of books that have the heading "Suicide." Mann makes the point that such books entered under the more general heading are effectively lost.³³ Unfortunately, too many people do not understand the principle, and so if a person finds a book under "Suicide" in which there is one chapter on suicidal behavior in youth, the person may not ever learn that there are whole books on the topic. This is a matter requiring not only correct cataloging but also understanding by those who help users with their searches.

There is concern on the part of some that even though the Core Record requires "the appropriate level of specificity," putting only one or two subjects in the Core Record will lead catalogers to use a more general term rather than follow the principle of specific entry. Elaine Svenonius and Dorothy McGarry discovered in their research evaluating subject headings assigned by non-LC catalogers that there has already been a problem long before the appearance of the Core Record. They found that over 50 percent of their sample had missing, incorrect, dated, or questionable headings, and the largest category of incorrect headings were the ones that were too broad.³⁴

One reason for the difficulty that we have giving specific entry has been identified by Miksa. Traditionally, libraries have opted to catalog physical units even when they contain multiple works. Thus, collections, proceedings, and "bound together"

items must be given a general heading appropriate to the contents of the whole physical item rather than several headings that are appropriate to the specific individual works. The tradition for commercial indexing enterprises has been that the basic unit is the work, not the package.³⁵ This may not be as much of a problem as we begin cataloging works from the Internet; we will be forced not to look at physical packaging and may be able to concentrate on cataloging individual works for which specific entry is more meaningful. Again, this is a matter that requires education if we see value in this principle and wish to maintain it.

CLASSIFICATION

A third area of controversy is classification. There is a school of thought that says that the main reason to classify is so that a physical item can be shelved in a place where it can be found again, and, oh, yes, incidentally, it is then with other items on the same subject. Those subscribing to this view see no reason to classify things that cannot be shelved—usually defined as anything that is not a book. Intner comments on the irony to be found in the fact that while some people are saying we no longer need subject headings and classification, others are asking for more subject access to fiction and nonprint materials.³⁶

Despite rumors to the contrary classification is far from dead. Both research and experiments have shown the usefulness of classification for enhanced subject access. Ray Larson has created a system in which subject clusters are created that include both LC Classification and Dewey Decimal Classification (DDC) numbers from a MARC record in the same cluster with subject headings and keywords from other parts of the record. The clusters make a package that allows several different kinds of subject access.³⁷ Karen Markey Drabenstott and Anh Demeyer have used the DDC numbers in records to enhance the subject term entry vocabulary by connecting the words associated with a particular number to the record in such a way that the words from the DDC can be used as search terms for the associated MARC records.³⁸ Barbara Tillett has suggested that classification could be the way links might be made among variant terminology of different thesauri.³⁹ Gorman has suggested that classification as a subject access tool be separated from classification as a place to shelve an item. A shorter number could be used for shelving purposes, while several longer numbers could cover the various subject aspects. In this way, the online catalog could "combine the best features of dictionary and classified catalogs."⁴⁰ [Actually, this might be stated: "best features of divided and classified catalogs." I have not seen an online catalog that gives one of the best features of a dictionary catalog: a searcher is seldom allowed to search for all works by or about a person in the same search unless one uses a very complicated Boolean search—which most users do not attempt.]

A recent *Newsweek* article explained how 3-D technology is hitting the Web. The description is of Cyberspace explorers wandering through whole virtual environments. "They will be able to cruise through representations of a city neighborhood or a mall, visiting shops simply by 'walking' down a corridor and clicking on the door."⁴¹ It is not hard to imagine in this scenario that one of the doors you might click on might be that of the "library" where you can enter various subject rooms with stacks of books. Going down the rows of stacks, you might locate a particular "book" you would like to "take down" from the shelf and peruse. However, in order to do anything like this, the "books" would have to be classified.

At the level of today's practice, a number of Internet sites are currently organized using the DDC as the organizing factor.⁴² In addition Mann has given numerous examples of reference questions that can only be answered by reference to the page and paragraph levels of books that can be found only by going to the classification in the stacks where such information might be found and then scanning the books located at that number. Sometimes such information is not even indexed in the back-of-the-book index. He makes the point that such information is not now, nor will it ever be, found in computer databases. Books that now exist in, and are still being added to, library collections are our cultural heritage, and while many of them will be digitized, most will not.⁴³

ONLINE PUBLIC ACCESS CATALOGS (OPACs)

Given that local collections of materials in libraries will continue to be a part of our world for the foreseeable future, it behooves us to be certain that the catalogs of those collections are useable and useful. Much of the current research in the area of subject access is directed toward this end. Larson, in his tracking of the use of the MELVYL online catalog system, discovered two phenomena that contributed most to reduced subject searching. Close to half of the time, subject searches retrieved nothing, and the remainder of the time, there were "too many" hits. Only about 12 percent of searches retrieved between 1 and 20 items.⁴⁴ Larson has identified several obvious causes of search failure: lack of knowledge of LCSH; misspelling or miskeying; use of singular or plural when the opposite would have matched; failure to predict the terms used to catalog the item.⁴⁵

"Too many" cannot be defined precisely, because it is relative to an individual's needs and to that individual's tolerance for scanning through screens of retrieved items. A number of researchers have found that users, faced with many retrieved items, often do not even *begin* browsing through the screens, and often when they do, they stop after the first screen or two. The problem with this is that in most online catalogs, retrieved items are arranged in "main entry" alphabetical order: so the best of the items may be one that is many screens into the listed retrievals, and may even be the last one listed! Buckland, Norgard, and Plaunt observe that, because each set of records retrieved has to be newly ordered to be displayed in that instance, there is no reason it has to be alphabetical. It could just as well be chronological, reverse chronological, by language, by proximity of holding library, by loan status, or by any number of other attributes; and it need not always be one of these but should be displayed by one or more of these chosen by the user.⁴⁶

Charles Hildreth advises that second generation online catalogs (i.e., what most libraries have currently) are so unacceptable that we should "Just say 'NO'" to them! He acknowledges that in many respects they are better than card catalogs, but he asserts that most reflect an incomplete conceptual model of information retrieval activity: "the model assumes the presence of a known specifiable information need (or subject topic) to start with."⁴⁷ Hildreth goes on to identify major shortcomings (and therefore major needs) of current online catalogs: they do not:

- Facilitate open-ended searching by following pre-established linkages between records; they do not;
- Assist the user with alternative formulations of search statements or with alternative search methods when initial searches fail;

- Lead from successful keywords to corresponding subject headings and class numbers;
- Provide sufficient information in the records (e.g., abstracts) to enable users to judge usefulness;
- Rank large retrieved sets in decreasing order of probable relevance.⁴⁸

Marcia Bates, too, discusses subject access in online catalogs and says that assistance for users is more essential than it was in the card catalog. She says that the online catalog is a "black box" to users. One cannot just look in and see what is there as one could with a card drawer; one has to put something in to get something out. In addition, implicit Boolean logic means that users need assistance with ideas for combining keywords from different parts of subject headings as well as from different indexes.⁴⁹

Various researchers have been working on ways to overcome the failures of today's OPACs. Karen Drabenstott's research attempts to find ways to route users' "zero-hit" queries through appropriate "search trees" so that the system does not just respond with a statement that there are no matches, but instead, performs other actions that are likely to produce some result. Search trees are also used to assist in narrowing a search on a term that is too broad, a search that often results in the number of retrievals being too large.⁵⁰ Later research shows that classification can be used to summarize the results of high-posted searches and thus make retrievals more manageable. Drabenstott finds that Dewey Decimal Classification is more useful for this than Library of Congress Classification.⁵¹

Tschera Connell has researched methodologies and knowledge used by experienced searchers. Academic librarians "thought aloud" as they performed subject searches. Connell collected information about their factual, experiential, and process knowledge, particularly concentrating on explanations that participants gave for their actions and observations. Connell observes that online catalogs too often "concentrate on mechanical aspects of searching (e.g., automatic truncation, spelling checkers), to the exclusion of intellectual aspects (e.g., structural relationships among records, controlled vocabulary maintenance)...."⁵² She has found that a relational structure must be present in the catalog if the user is to exploit the materials represented in the database.

Larson's research investigates an OPAC structure called "classification clustering." Such a cluster includes classification numbers and their meanings, subject headings, titles, and other words from a MARC record in a new record that is the first level searched.⁵³ Other experiments to improve retrieval involve enhancing records with tables of contents, more classification numbers, and more subject headings; reducing search terms to single root words (e.g., "librar" for library, libraries, etc.); and use of phonetic indexes.⁵⁴ The difficulty is that all these help to retrieve more, and as we have already seen, much of the time users retrieve too much. So there are also experiments to assist with the problem of information overload.

One experiment that deals with retrieval of too many records is an "intelligent" front end called OASIS added to the University of California's MELVYL online catalog. Buckland says that "The design ideal is that regardless of the query and regardless of the size and contents of the file, each search will ordinarily yield a small set of records, those most closely matching the searcher's preferences."⁵⁵ Users can set their own preferences, but the default assumes that users want recent materials in

English held locally. This process is called "filtering." So, for example, a subject keyword search for "Dresden" that yields 440 hits can be filtered to find that there are 9 books in English published since 1980 that are held on the Berkeley campus. These can be looked at first. OASIS also can "summarize subjects," a command that causes the system to go through the retrieved list and produce a frequency-ordered list of the subject headings that occur on the records. The user can then choose any one and have it reissued as a new search.

Another way of dealing with overload is to have the system calculate a probability that each item retrieved is an item that the user would want to see. In Larson's research the clusters can be weighted so that probabilistic retrieval techniques can be combined with cluster searching to produce ranked output.⁵⁶ It is possible to compare a user's natural language query with the probable relevance of each MARC record based upon the occurrence of words in the cluster.

LC SUBJECT HEADINGS

If it is true that we will continue to maintain online catalogs and that in these catalogs there is a need for controlled vocabulary, we will probably be using Library of Congress Subject Headings (LCSH) for the foreseeable future. A few years ago there was considerable discussion of coming up with a new system, or completely overhauling LCSH, or creating a subject cataloging code. However, each of these alternatives would require considerable resources and commitment on the part of the library and information science community. Drabenstott and Vizine-Goetz identify six reasons that LCSH is here to stay:

1. Public catalog users with varying levels of subject knowledge can find materials on the topics they seek using LCSH.
2. In terms of subject retrieval, a better system has not been devised, nor will be in the near future.
3. Online systems have capitalized on some desirable features of the LCSH system.
4. Online systems have overcome some limitations of the LCSH system.
5. Improvements can be made to LCSH.
6. Tens of millions of bibliographic records are available to the users of online bibliographic systems through LCSH.⁵⁷

Work has proceeded on making improvements to LCSH, and progress is being made. Since 1985 when LC revised its rules for the reference structure to accompany new headings, many more clearly delineated hierarchical paths are emerging. In 1991 LC held a conference to invite suggestions on how they might improve the existing system of subdivisions. As a result pattern lists of subdivisions have been simplified; some subdivisions have been cancelled; some have been combined with other almost equivalent subdivisions; some have been changed to more current forms; some have been replaced by more logical phrases (e.g., "Railroads—Stations" changed to "Railroad stations"); date subdivisions have been changed to relate to the coverage of contents of an item rather than to its date of issue.

Three of the recommendations of the subject subdivision conference warrant special mention. The first recommendation included the provision that subdivisions should always appear in the order: "topical, geographic, chronological, form."⁵⁸ After considerable study, a subcommittee of the ALA/ALCTS/CCS Subject Analysis Committee (SAC) found that it could not recommend implementation of this across the board. They found

that in the areas of History, Art, Literature, and Music, there would be too many times when the practice would make the subject headings lose their current meaning, and therefore would not be in the best interests of users. However, it was noted that in most other areas, the recommendation could be implemented without ill effect, and they recommended that LC follow the recommendation whenever possible in order to improve consistency and predictability. LC has, on a case-by-case basis, changed many topical subdivisions to allow them to be subdivided by place. Thus, place no longer has to precede those subdivisions and can come in the position recommended by the subdivisions conference.

It should be noted that in the process of examining the problem of the order of subdivisions, one study found that users understand the meaning of current subject heading strings only about 40% of the time and that they would understand the strings in the recommended order only about 32% of the time. The thing the researchers found that increased understanding was shorter strings. The authors therefore recommended reducing the number of subdivisions per string and reducing the number of words in individual subject heading subdivisions.⁵⁹

The second recommendation of the subject subdivisions conference that bears special mention is the recommendation that LC should investigate implementing a separate subfield code for form subdivisions. A new code (subfield \$v) was approved for addition to the USMARC formats at the 1995 Midwinter ALA Conference in Philadelphia. SAC, LC, and others are proceeding with some preliminary stages necessary before implementation of the new code.

The third recommendation that bears mention here is the recommendation that authority records be created for combinations of topical headings with topical subdivisions. LC has been working with staff from OCLC's Office of Research to develop a base file of such strings derived from LC bibliographic records in the OCLC database. In an OCLC symposium on the future of authority control Carol Mandel discusses what a real LCSH-based subject authority file might look like and what might be its benefits. She states that it could speed the cataloging process with some automatic assistance (e.g., one-to-one references for synonyms); could provide quick access to valid forms; could require less proofreading; and could have the potential for more automatic generation of references, links from classification numbers to headings, and links from bibliographic records to authorized forms of headings. For users it could assist with typographical errors; has the potential to correct user input with algorithms; and has the potential for better browsing displays of subject terms. However, we have to be willing to make a one-time large investment in order to have ongoing savings from machine support.⁶⁰

For the last couple of years, LC has strongly encouraged other libraries to submit proposals for adding new headings/concepts to LCSH, changing their longstanding practice of waiting to add new concepts until they were encountered in LC's own cataloging. The result is the addition in a controlled manner of new concepts needed by many catalogers outside of LC. Forms for suggesting new headings are available on the Internet or through "gopher" to LC MARVEL. LC has also expanded its programs by which participating libraries actually create new subject headings and make changes to existing headings in the growing "national authority file."

CONCLUSION

This article began with quoting some persons who have questioned the need for subject access in library catalogs. We then showed that much recent writing and research has demonstrated the need of many users for accurate and reliable subject approaches. Subject organization of the Internet will, in the words of Michael Gorman, be a "disappointment to those who have made a vocation of surfing the net as an alternative to living."⁶¹ For most of us, though, subject organization will make life ever so much easier. It is another example of the issue of the time of one cataloger vs. the time of multiple users.

Will we continue to catalog important intellectual entities for posterity, as well as for current users? Carol Mandel points out that "The library catalog record, with its summarized description and controlled [subject] vocabulary is an extraordinarily efficient and effective representation of text for retrieval."⁶² Without the subject vocabulary, the catalog record identifies a known item but gives no clue to content unless the title has content-descriptive words. The budget vise squeezing librarians, identified by Carlen Ruschoff,⁶³ may cause some to be tempted to cut out subject analysis as a way to cope. But the evidence suggests that what is needed for subject analysis, as it is needed for the rest of cataloging, is innovation: new ways of separating that which is in need of less subject analysis from that which is in need of more subject analysis; better education of reference librarians as to what subject headings and classification can do for them and for their clients; further national-level cooperation in the creation of subject access; and implementation of the techniques that research has shown will greatly reduce frustration from retrieving nothing or retrieving too much.

REFERENCES

1. Ruth French Strout, "The Development of the Catalog and Cataloging Codes," *Library Quarterly* 26 (October 1956): 267-268.
2. Karen Markey, *Subject Searching in Library Catalogs: Before and after the Introduction of Online Catalog*, (Dublin, Ohio: OCLC Online Computer Library Center, 1984).
3. J. Bradford Young, personal communication, 1994.
4. Thomas Mann, "Cataloging and Reference Work," in *Crisis in Subject Cataloging and Retrieval* (ALCTS), recorded at the American Library Association Annual Conference (Chicago, June 1995), audiotape.
5. Ray R. Larson, "The Decline of Subject Searching: Long-Term Trends and Patterns of Index Use in an Online Catalog," *Journal of the American Society for Information Science* 42 (April 1991):197-215.
6. Peter S. Graham, "Quality in Cataloging: Making Distinctions," *Journal of Academic Librarianship* 16 (September 1990): 218.
7. Larson, "The Decline of Subject Searching."
8. Larson, Ray R. "Intelligent Information Retrieval Using Controlled Vocabulary in MARC Records," in *Artificial Intelligence and Authority Control: Together at Last?* recorded at the American Library Association Annual Conference (Miami Beach, June 25, 1994), audiotape.
9. Larson, "The Decline of Subject Searching," p. 207
10. Michael Gorman, "The Cost and Value of Organized Subject Access," in *Crisis in Subject Cataloging and Retrieval* (ALCTS), recorded at the American Library Association Annual Conference (Chicago, June 1995), audiotape.
11. Robert L. Jacobson, "Taming the Internet," *The Chronicle of Higher Education* (April 21, 1995), p. A29.
12. Stuart C. A. Whitwell, "Redeeming Value," *American Libraries* 26 (June 1995): 555.
13. James Beach, [Presentation] in *Authority Control in an Integrated Archive, Library, and Museum Information Environment* (LITA/ALCTS), recorded at the American Library Association Annual Conference (Chicago, June 25, 1995), audiotape.
14. Francis L. Miksa, "Bibliographic Control Traditions and Subject Access in Library Catalogs," in *Crisis in Subject Cataloging and Retrieval* (ALCTS), recorded at the American Library Association Annual Conference (Chicago, June 1995), audiotape.
15. Karen G. Schneider, "Welcome to Karen's Kitchen!" 1995, (in Her Home Page: <http://nile.intac.com:80/~kgs/>)
16. Ibid.
17. Jacobson, "Taming the Internet," p. A29.
18. C. Olivia Frost & Joseph Janes, "An Empirical Test of Gopher Searching Using Three Organization Schemes," in *Annual Review of OCLC Research 1994* (Dublin, Ohio: OCLC Online Computer Library Center, 1995), pp. 54-56.
19. Robert Pool, "Turning an Info-Glut Into a Library," *Science* 266 (October 7, 1994): 20-22.
20. "OCLC NetFirst to Improve End-User Access to Internet Resources," press release, April 7, 1995.
21. Willy Cromwell, "The Core Record: A New Bibliographic Standard," *Library Resources & Technical Services* 38 (October 1994): 415-424.
22. Richard Giordano, "The Documentation of Electronic Texts Using Text Encoding Initiative Headers: An Introduction," *Library Resources & Technical Services* 38 (October 1994): 389-401.
23. *OCLC/NCSA Metadata Workshop Report, 1995* (found on the Internet at: http://www.oclc.org:5046/conferences/metadata/dublin_core_report.htm)
24. Thomas Mann, *Doing Research at the Library of Congress: A Guide to Subject Searching in a Closed Stacks Library* (Washington, D.C.: Humanities and Social Sciences Division, Library of Congress, 1994), p. 7.
25. F. W. Lancaster, *Vocabulary Control for Information Retrieval*, 2nd ed. (Arlington, VA: Information Resources Press, 1986), p. 6.
26. Lancaster, *Vocabulary Control for Information Retrieval*, p. 8.
27. C. P. R. Dubois, "Free Text vs. Controlled Vocabulary: A Reassessment," *Online Review* 11 (1987): 249.
28. Ibid., p. 243.
29. Cromwell, "The Core Record: A New Bibliographic Standard," p. 420.
30. Charles Ammi Cutter, *Rules for a Dictionary Catalog*, 4th ed. (Washington, D.C.: GPO, 1904), p. 66-67.
31. David Judson Haykin, *Subject Headings: A Practical Guide* (Washington, D.C.: GPO, 1951), p. 9.
32. Mann, *Doing Research at the Library of Congress*, p. 9.
33. Mann, "Cataloging and Reference Work."
34. Elaine Svenonius & Dorothy McGarry, "Objectivity in Evaluating Subject Heading Assignment," *Cataloging & Classification Quarterly* 16 (1993): 25.
35. Francis L. Miksa, "Bibliographic Control Traditions and Subject Access in Library Catalogs," in *Crisis in Subject Cataloging and Retrieval* (ALCTS), recorded at the American Library Association Annual Conference (Chicago, June, 1995), audiotape.
36. Sheila S. Intner, "Subject Access Education—Oxymoron or Obligation?" in *Crisis in Subject Cataloging and Retrieval* (ALCTS), recorded at the American Library Association Annual Conference (Chicago, June, 1995), audiotape.
37. Ray R. Larson, "Classification Clustering, Probabilistic Information Retrieval, and the Online Catalog," *Library Quarterly* 61 (1991): 133-173.
38. Karen Markey & Anh N. Demeyer, *Dewey Decimal Classification Online Project: Evolution of a Library Schedule and Index Integrated into Subject Searching Capabilities of an Online Catalog* (Dublin, OH: OCLC, Office of Research, 1986).
39. Barbara Tillett, [Presentation] in *Authority Control in an Integrated Archive, Library, and Museum Information Environment* (LITA/ALCTS), recorded at the American Library Association Annual Conference, (Chicago, June, 1995), audiotape.
40. Gorman, "The Cost and Value of Organized Subject Access."

41. Michael Meyer. "Surfing the Internet in 3-D." *Newsweek* (May 15, 1995), 68.
42. OCLC Forest Press. "Internet Resources Classified by Dewey." 1995 (found on the Internet at: <http://www.oclc.org/oclc/fp/ddc/resource.htm>)
43. Mann, "Cataloging and Reference Work."
44. Larson, "The Decline of Subject Searching," pp. 207-210.
45. Larson, "Intelligent Information Retrieval Using Controlled Vocabulary in MARC Records."
46. Michael K. Buckland, Barbara A. Norgard, & Christian Plaut, "Filing, Filtering and the First Few Found," *Information Technology and Libraries* 12, E³ (Sept. 1993): 313.
47. Charles R. Hildreth, "Advancing toward the E3OPAC: The Imperative and The Path," *Think Tank on the Present and Future of the Online Catalog: Proceedings, RASD Occasional Papers*, Number 9, ALA Midwinter Meeting, January 11-12, 1991, Chicago, Illinois, Sponsored by the RASD Catalog Use Committee, edited by Noelle Van Pulis (Chicago: ALA/RASD, 1991), p. 23.
48. *Ibid.*, pp. 30-31.
49. Marcia J. Bates, "Rethinking Subject Cataloging in the Online Environment," *Library Resources & Technical Services* 33 (October 1989): 404-405.
50. Karen Markey Drabenstott & Diane Vizine-Goetz, *Using Subject Headings for Online Retrieval: Theory, Practice, and Potential* (San Diego: Academic Press, 1994), pp. 144-145.
51. Karen M. Drabenstott (with research assistance of Celeste M. Burman & Marjorie S. Weller), *Enhancing a New Design for Subject Access to Online Catalogs* (Ann Arbor, MI: School of Information and Library Studies, The University of Michigan, 1994), pp. 144-145.
52. Tscherna Harkness Connell, "Subject Searching in Online Catalogs: Metaknowledge Used by Experienced Searchers," *Journal of the American Society for Information Science* 46 (August 1995): 516.
53. Larson, "Classification Clustering, Probabilistic Information Retrieval, and the Online Catalog."
54. Larson, "Intelligent Information Retrieval Using Controlled Vocabulary in MARC Records."
55. Buckland et al., "Filing, Filtering and the First Few Found," p. 314.
56. Larson, "Classification Clustering, Probabilistic Information Retrieval, and the Online Catalog."
57. Drabenstott & Vizine-Goetz, *Using Subject Headings for Online Retrieval*, p. 331.
58. Library of Congress, *The Future of Subdivisions in the Library of Congress Subject Headings System: Report from the Subject Subdivisions Conference*, edited by Martha O'Hara Conway (Washington, D.C.: Library of Congress, 1992).
59. Lori Franz, John Powell, Suzann Jude, & Karen M. Drabenstott. "End-User Understanding of Subdivided Subject Headings," *Library Resources & Technical Services* 38 (July 1994): 213-226.
60. Carol Mandel, "Change and Continuity in Subject Authority Control," in *The Future Is Now: Reconciling Change and Continuity in Authority Control: Proceedings of the OCLC Symposium*, ALA Annual Conference, June 23, 1995 (Dublin, OH: OCLC Online Computer Library Center, 1995), videotape.
61. Gorman. "The Cost and Value of Organized Subject Access."
62. Mandel, "Change and Continuity in Subject Authority Control."
63. Carlen Ruschoff, "Cataloging's Prospects: Responding to Austerity with Innovation," *Journal of Academic Librarianship* 21 (January 1995): 51-57.

Library Benchmarking: Old Wine in New Bottles?

by Sarah M. Pritchard

When the term "benchmarks" first crept into the library and higher education lexicons, it was with such an aura of novelty and efficacy that I thought it must be a complicated and technical approach requiring special study. And then I realized that I already gathered and used benchmarks, I just did not call them that. Moreover, the flood of recent articles has not given us a new formula for deciding what is a "good" library. This is an old problem, and though it is one very much worth tackling, I am afraid we have not found a new solution.

Benchmarks have caught on with the advent of total quality management (TQM) programs, though the origins of the term date earlier to engineering and scientific industry work. It seems to have two contrasting meanings: either something to work towards, or something to mark how far you have come. It can be at the same time measurement against standards, against one's own prior progress, or against wherever a peer or competitor happens to be. The fact remains, we are simply talking about measurement. In this profession we have been trying, with varying degrees of success, to measure what we do for at least the last century. Benchmarks are nothing more than a management- or production-oriented way of talking about statistics. Perhaps it is even a more accurate word for what we have always done. Statistics include more sophisticated analyses than those in a typical compilation of what we call library statistics, most of which are just data elements, singly or in ratios—in a word, benchmarks.

Benchmarks are not synonymous with TQM; they are but one of the recommended tools for implementing such a system. The point is not what we call these measures, but how we define them and use them. Statistics, benchmarks, and any kind of data are always just tools; they are not an end in and of themselves. The same data can be used and reused in a variety of larger contexts, not all predictable or even appropriate. The data are the answer to a *prior question*; they have no meaning in the abstract. One could even posit a "quantum theory" of library statistics, that they do not exist until we go out and collect them, and it is the act of collecting them that defines them.

Developing appropriate benchmarks is part of a process of evaluation, a long tradition in library management. While the language and some of the processes may be new, I have not, unfortunately, seen any magic units or measures in the literature of "quality" or assessment that will give us quick relief from the hard practical and conceptual challenges that the evaluation enterprise continues to present. Let us demystify this topic a bit, and get back to the fundamental problem of library measurement. This is an attempt to clarify definitions, outline some ways in which we have developed benchmarks already, and sketch the importance of this for information services management in the context of contemporary higher education.

DEFINING BENCHMARKS AND STATISTICS

These terms are loosely applied to a range of numerical charts and data, some of which are actually quite simple. We may be both over-intimidated and over-credulous by our too casual

Sarah M. Pritchard is Director of Libraries, Smith College, Northampton, Massachusetts 01063.



ISSN 0099-1333
Vol. 21, No. 6

THE JOURNAL OF ACADEMIC LIBRARIANSHIP

November 1995

EDITOR-IN-CHIEF

Peter Hernon
Simmons College, Boston

EDITOR

Charles Martell
California State University,
Sacramento

FOUNDING EDITOR

Richard M. Dougherty

BOOK REVIEW EDITORS

Peter Hernon & Christy Zlatos

JAL GUIDE EDITORS

Ellen Altman
Allan Pratt

EDITORIAL BOARD

Don L. Bosseau
San Diego State University

Nicholas C. Burckel
Marquette University

Susan M. Campbell
York College of Pennsylvania

Robert E. Dugan
Wesley College

Donald G. Frank
Georgia Institute of Technology

William Gosling
University of Michigan

Roma Harris
University of Western Ontario

Eddy Hogan
California State University,
Sacramento

Cheryl LaGuardia
Harvard University

Laurie S. Linsley
Seminole Community College

Charles Lowry
Carnegie Mellon University

Susan K. Martin
Georgetown University

Cheryl Metoyer-Duran
University of California, Riverside

Sarah M. Pritchard
Smith College

Carlen Ruschoff
Georgetown University

Allen Smith
Simmons College

Helen H. Spalding
University of Missouri-
Kansas City

NOTICE: This material may be
protected by Copyright Law
(Title 17 U.S. Code)

Editorial: Libraries As Places in a Time of Rapid Change 425
by Peter Hernon

ARTICLES

**The Micropolitics of Budgeting in Universities:
Lessons for Library Administrators**
by Douglas G. Birdsall 427

Stakeholder Perceptions of University Library Effectiveness
by Rowena J. Cullen and Philip J. Calvert 438

**Favorable and Unfavorable Book Reviews:
A Quantitative Study**
by Robert Greene and Charles Spornick 449

**Perceptions vs. Use: Comparing Faculty Evaluations of
Journal Titles with Faculty and Student Usage**
by Kathleen E. Joswick and Jeanne Koekoek Stierman 454

**Subsidizing Unmediated Document Delivery:
Current Models and a Case Study**
by Minna Sellers and Joan Beam 459

NOTICE: This material may be
protected by Copyright Law
(Title 17 U.S. Code)

The Journal of Academic Librarianship (ISSN 0099-1333) is published bimonthly beginning January 1994 by JAI Press Inc., 55 Old Post Rd. No. 2, P.O. Box 1678, Greenwich, CT 06836-1678. Subscription rates: \$50 per year to individuals; \$125 per year to institutions. Foreign subscribers add \$20. All remittances must be in U.S. dollars. Single copies and selected back issues are available at \$15 each. Second class postage paid in Ann Arbor, Michigan. Address all advertising correspondence to: Miranda Garside, JAI Press Inc., 2621 O. K. and Drive, Kalamazoo, MI 49008, (616) 385-3306, Fax: (616) 385-3307.

The Journal of Academic Librarianship is indexed in *Aslib Current Awareness Abstracts*, *Current Contents*, *Current Index to Journals in Education (CIJE)*, *Education Index*, *Information Science Abstracts*, *Library & Information Science Abstracts*, *Library Literature*, and *Social Sciences Citation Index*. It is also included in *Academic Abstracts*, *Articles First (OCLC Database)*, *CARL UnCover*, *Contents First (OCLC Database)*, and *Current Awareness Abstracts*. Book reviews are indexed in *Book Review Index*.

Postmaster: Send address changes to JAI Press Inc., 55 Old Post Rd. No. 2, P.O. Box 1678, Greenwich, CT 06836-1678 © 1995 by JAI Press Inc., except for works by those employed by the U.S. government. All rights of reproduction in whole or in part in any language without written permission is prohibited.