Improved Browsable Displays for Online Subject Access

Mia Massicotte,
Systems Librarian, Concordia University Libraries, Montreal, Quebec

Published in: *Information Technology and Libraries* 7.4 (1988): 373-380

One of the obstacles in the way of effective subject access in OPACs (online public access catalogs) is undue length of a browsable index display under a specified search term. The complexities associated with online subject access have been firmly documented in the literature, notably by Cochrane, Markey, Mandel and Herschman, and others.[1] Well-founded criticism of *LCSH (Library of Congress Subject Headings*) and LC practice need to be addressed.[2] Even the unlikely event of a major restructuring of *LCSH*, however, cannot solve the myriad problems facing effective subject access. We must undertake a more practical problem-solving approach, and quickly, if we are to remove some of the hurdles facing our OPAC users today. By concentrating solely on the browsable index display, we may move towards eliminating some of the difficulties currently encountered during subject searches while at the same time exploiting, to the user's advantage, the existing large body of subject data resident in our bibliographic files.

While browsing a subject index is a desirable feature in an OPAC, often the display under a specified search term is extremely lengthy due to the number of index entries clustered under that term in alphabetical proximity. Many OPACs display a particular subject index term on one line, followed by a list of each unique subdivision arranged alphabetically on successive lines of the display (see figure 1).

Islam
Islam (the Arabic word)
Islam -- Addresses, essays, lectures
Islam -- Africa
Islam -- Africa -- Addresses, essays, lectures
Islam -- Africa -- Bibliography
Islam -- Africa -- Congresses
Islam -- Africa, East
Islam -- Africa -- History
Islam -- Africa, North -- History
Islam -- Africa, Northwest -- History
Islam -- Africa, Subsaharan -- Bibliography
Islam -- Africa, West
Islam -- Africa, West -- History -- Addresses, essays, lectures
Islam -- Algeria
Islam -- Algeria -- Addresses, essays, lectures
.
.
Islam -- Collected works
.
.
Islam -- Dictionaries

Fig.1

[1] Pauline Atherton Cochrane, *Redesign of Catalogs and Indexes for Improved Online Subject Access: Selected Papers of Pauline A. Cochrane* (Phoenix, Ariz.: Oryx Pr., 1985); Karen Markey, *Subject searching in Library Catalogs: Before and after the Introduction of Online Catalogs* (Dublin, Ohio: OCLC Online computer Library Center, 1984); Karen Markey, "Users and the Online catalog: subject access problems," in *The Impact of Online Catalogs*, ed. Joseph R. Matthews (New York: Neal-Schuman, 1986), p.35-69; Carol A. Mandel and Judith Herschman, *Subject Access in the Online Catalog: a report prepared for the Council on Library Resources* (Washington, D.C.: Council on Library Resources, 1981).

[2] Monika Kirtland and Pauline Cochrane, "Critical views of LCSH--Library of Congress Subject Headings: A bibliographic and bibliometric essay," *Cataloging & Classification Quarterly* 1, no.2/3:71-94 (1982); William Mischo, "Library of Congress Subject Headings: a review of the problems, and prospects for improved subject access," *Cataloging & Classification Quarterly* 1, no.2/3:105-24 (1982).

Interspersed in this list are other types of headings (e.g. Inverted word order) that machine filing embeds as a result of the alphabetical sorting routine. For many users, browsing the subject index has meant combing through large numbers of subject entries that have been machine indexed in alphabetical order.[3]

In a lengthy index display of a specified search term--consider anything over one screen lengthy--the user is generally assaulted by all kinds of alphabetically arranged subdivisions. Prompting the user to limit the search by a variety of available techniques, such as Boolean operators, word proximity, and the like, may not be viewed as particularly helpful suggestions, no matter in what friendly language they are couched. The user in many cases does not know what is required in order to be more specific. Add to this the copiously documented problems of lead-in vocabulary, level of indexing of the documents themselves, spelling, syntax, command semantics and protocol, and so forth, and one of which can place insurmountable obstacles in the user's path, and we can easily conclude that we are far from providing users with an efficient service.

Description of the Problem

The problem we are faced with is undue display length of a browse list under a given search term. Scanning an alphabetically arranged list of indexed terms is simply not sufficient to assist the user. New character strings are added daily to the index, and the indexes will continue to expand at an ever-increasing rate. This factor alone will eventually make browsing the alphabetical index less and less viable as a method of searching. If we begin to index our documents more exhaustively, for example, by adding more subject headings per record as recommended by the CLR-sponsored subject access meeting, we will arrive at this stage even earlier.[4]

Does this mean that browsing large subject indexes cannot be an effective method of subject searching? Emphatically no. However, allowing the option to browse through the subject index does not absolve system designers of the responsibility to organize the index display for maximum gain. Take the example of a user who realizes immediately upon browsing a subject term that geographical aspects of the subject are irrelevant to the search. Nevertheless, geographic subdivisions are dutifully displayed, dispersed throughout the alphabetical array of subdivisions. Successive screens are cluttered with this information, and the user is forced to browse this list to pick out potentially meaningful subdivisions. The subdivisions themselves (including the generally unhelpful types: "-- Addresses, essays, lectures"; "-- Miscellanea"; and so forth) are not grouped in a conceptually fashion but simply filed alphabetically. Given the existing structure of *LCSH*, a way must be found to manipulate the index itself, and the new terms continually added to it, in order to arrive at a coherent and succinct display of the subject index terms in our files.

Getting around the problem

In the same way that subject headings attempt to provide a conceptual framework in which to collocate materials (however faulty that framework is), so too can subject subdivisions be grouped to provide further definition of the original concept.

An online catalog should orient the user to what is contained in the database and provide a conceptual framework within which a search may be freely and easily conducted, allowing serendipity to come into full play. If such a conceptual framework were presented to the user, despite how broad a topic was

---

[3] Pauline A. Cochrane, *Improving LCSH for Use in Online Catalogs: exercises for self-help with a selection of background readings* (Littleton, Colo.: Libraries Unlimited, 1986), p.59-62.

[4] Keith W. Russell, ed., *Subject Access: Report of a meeting sponsored by the Council on Library Resources, Dublin, Ohio, June 7-9, 1982* (Washington, D.C.: Council on Library Resources, 1982), p.71.

searched, and furthermore, if it were presented on one screen, a user could place as-yet-undefined interests into that framework without having further obstacles placed directly in the search path

What is required, then, is a sort of meta-index: a conceptual, rather than alphabetical, index of the individual index entries grouped under a particular search term or stem. Let us take the list of main, unsubdivided LC subject headings as a given. As pointed out earlier, these already reside in millions of bibliographic records. They are further subarranged A through Z by numerous subdivisions, which, in turn, may likewise be subdivided. Users cannot be expected to intuit all the possible subdivisions we have indexed our documents under, but in reality, this is what we are tacitly expecting them to do.[5] Neither will a user be readily able to detect all subdivisions that may potentially be relevant to the search, however specific or general the topic is formulated in the user's mind. To date, we have relied heavily on the persistence of the user to select or reject our self-imposed and somewhat arbitrary distinctions.

To help users define their interests in a way that will coincide with the system's indexing terminology, we must begin to construct a conceptual framework that will be immediately and intuitively accessible to them. If we look at geographical subdivisions, for example, and use it to provide us with a working model, we may begin to examine how the existing MARC record structure could be combined with better system design to provide the user with a mental "picture" of the database.

All subdivisions in MARC subject headings (6XX fields) are preceded by a subfield code. A routine could be employed to vastly reduce the size of index displays by substituting a general message for all records subdivided by a particular subfield code. In the case of the z subfield code for geographical subdivision, for example, the substituted text in the index display could be a phrase such as "subdivided by geographical areas".

A recently conducted browse of the Utlas database index under the term "Islam and politics" can better illustrate this point. While Utlas is a bibliographic utility, not an OPAC, the enormous size of its browsable indexes can provide us with some idea of how large files behave in general.

A browse of the Utlas index under the heading "Islam and politics" (see figure 2) retrieved ninety-six index entries – far too many to view on three or even four successive screens (the maximum single-screen display in the Utlas browsable index is twenty lines). A closer look at the individual index entries revealed that of the ninety-six index entries, only five entries were for titles. Of the remaining ninety-one entries for subject headings, eighty-one happened to have geographic subdivisions.

Browse of the Utlas index

1 Islam and politics
2 Islam and politics Addresses essays lectures
3 Islam and politics Adresses essays lectures
4 Islam and politics Afghanistan
5 Islam and politics Africa
6 Islam and politics Africa North
7 Islam and politics Africa North Addresses essays lectures
8 Islam and politics Algeria
.
.
48 Islam and politics Malaysia
.
.
96 Islam and politics Turkey History 20th century Congresses
97 Islam and power

Fig. 2. Browse of the Utlas index

[5] The same is of course true for the main term or terms under which we have chosen to index the document. This is an old dilemma, and one we are far from resolving.

For purposes of the index display, if we were to compress the eighty-one records where a z geographical subfield code is present by substituting a textual message, we might be presented with a screen that looks something like figure 3.

1 Islam and politics
2 Islam and politics – Addresses, essays, lectures
3 Islam and politics – Congresses
4 Islam and politics – Early works to 1800
5 Islam and politics – History
6 Islam and politics – Juvenile films
7 Islam and politics – Miscellanea
8 Islam and politics – Periodicals
9 Islam and politics – SUBDIVIDED BY GEOGRAPHICAL AREA, E.G. AFGHANISTAN

Fig. 3. Compression and substitution of textual message

Though the display is far from optimal at this point, it represents a step in the right direction in terms of the display and management of the data contained in the subject headings. Certainly much of the "noise" present in the original display has been eliminated. It is evident at the outset that a conceptual framework is operative, given the textual message appearing on the last line of the display. Furthermore, we can now see that other groups of concepts have clearly emerged, such as the format "Juvenile films". While at first it might appear that the screen is too simple, this is precisely what we should be aiming toward. The sample screen shown in figure 3 is a condensation of what is represented in the database under that subject heading, and a clear one at that.

If we extend this same method of compressing index entries and substituting textual messages for the chronological subdivisions that are preceded by the y subfield in the MARC record, we can achieve similar results in the reduction of the size of the display while simultaneously providing the user with a certain context. Geographical and historical subdivisions may now be easily rejected or pursued, although the user may not have known to specify or exclude them in the initial search strategy. By presenting subcategories, we present options to the user, thereby automatically incorporating some of the benefits of explicit Boolean searching.

The Challenge of the x subfield

With some caution, we may now approach the problem of how to handle the hundreds of subdivisions present in the x subfield. This subfield presents us with the greatest challenge for improving browsable subject displays since it has served as a catch-all for subdivisions that are neither geographical nor chronological in scope. In it are incorporated form and topical subdivisions, which are free-floating (not linked to a particular heading, such as "Congresses"), as well as specialized topical subdivisions (such as "Automobile motors – Fuel injection"). Since many topics are subdivided because no other suitable terminology exists (for a variety of reasons), the non-free-floating subdivisions stand to benefit the most from a hierarchical restructuring of *LCSH* in combination with computer-manipulated phrase-rotation and enhanced cross-references to improve lead-in vocabulary. Quite apart from the topical subdivisions, however, the list of free-floating subdivisions is a logical starting place for considering alternative methods of improving browsable displays.

A look at the list of free-floating subdivisions reveals a redundancy of related terms that can also be grouped into larger categories. For example, seventeen unique, free-floating subdivisions listed in the *LC*

Guide to Subdivision Practice represent graphic or illustrative aspects of a subject.[6] Similarly, there are subdivisions representing legal, technical, mathematical, scientific, economic, sociological, and political aspects of a subject. There are, of course, different ways in which the categories could be established; but, optimally, the first level of display should generate broad categories that can be enumerated on a single screen.

It would not be difficult to begin by examining the list of free-floating subdivisions to establish general categories under which they themselves can be grouped. By conceptually linking related terms, a table of existing subdivisions could be built, as long as we can specify what text should be mapped into the online display. Since the list of free-floating subdivisions is finite, we could specify, for example, that under any given main heading, "when two or more of the following seventeen terms appears in the x subfield, substitute the text 'Graphic or illustrative aspects' in the index display." The same can be done for concepts relating to legal aspects, mathematical aspects, and so forth. We needn't worry if a particular free-floating subdivision may be interpreted in more than one way because the tables do not need to be mutually exclusive. The same term might appear both in a table of "economic aspects" as well as in a table for "political aspects" if such is the intent of the *LCSH* scope note governing how that subdivision is to be used. Using this as a point of departure, broad conceptual groups could help reduce the size of a browsable display under a specific index terms to a maximum of one screen at the first level of display.

What we have is a type of menu approach where users further select their areas of interest. Unlike a true menu-driven system, however, the design does not actually depend on a hierarchical tree structure but in some respects behaves as though it does. Flexibility is not restricted by lengthy search paths that can disorient the user, and the user may employ other conventional means of limiting the search at any desired point. By focusing on a very small group of concepts initially, the user picks up cues as to how a topic is subarranged in the database. The user does not have to know beforehand which possible categories are available and is also spared from having to make sense of what may appear to be arbitrary distinctions. It also allows the user to invisibly bridge some of the terminological and indexing obstacles that previously stood in the way.

As the user gets conceptually closer to an area of interest, other traditional techniques for limiting search may *now* be gainfully employed. Suppose the user has determined that geographical areas are of interest and selects that category. The system might respond by automatically displaying the possibly geographical areas available for selection. If that is itself would be unmanageable (due, for example, to exceeding the single screen limit on the size of index displays), the system may prompt the user to enter the geographical area.

Thus, we have somewhat skirted the problem of requiring the user at the early stages of inquiry to have prior knowledge of what is available in the universe of the database. We have not required the user to hunt haphazardly through successive screen of subdivisions. For the user who has, by accident or design, selected a particular subdivision, we have minimized the danger of overlooking other relevant subdivisions -- subdivisions that were not made meaningful due to terminological ambiguity, obscurity, or irrelevant context. By conceptually linking related subdivisions in the index display, we could help users visualize the categories that reflect the subject *content* of the database. With a conceptual framework, users could more effectively and easily find their way around.

Possible methods available to us

Using compression techniques in combination with tables in which the finite set of existing character strings in the x subfield can be mapped opens up new possibilities for more effective system design. Each unique, main subject heading should appear only once in the general browsable display, no matter how many subdivisions are linked to it. In the same way that browsing a name authority file should reveal

[6] Library of Congress, Subject Cataloging Division, *Library of Congress Subject Headings: A guide to subdivision practice* (Washington, D.C.: LC, 1981).

only one occurrence for each unique name, so should the subject index be similarly constructed. The general index should reveal a list of main subject terms, and term selection should further reveal levels of subcategories suitably arranged in a conceptual (not necessarily alphabetical) fashion.

We may consider methods of manipulating existing subject heading data for more effective browsable displays, keeping in mind some characteristics peculiar to present and past subdivision practice.

1. Under any main subject heading, match and compress each unique sorting element within the geographical subfield (for example, all occurrences identified by the same character string, e.g., Africa) so that each unique geographic area occupies only one line in the appropriate level of browsable display. This would eliminate the problem of compound headings or headings with inverted word order splitting a file (see entries above and below the heading "Islam -- Africa, East" in figure 1). It would also cluster all matching subject headings, even when further subdivisions are present. The second level of display in figure 4 gives some idea of how this particular method could operate. A hierarchy of subject heading index display would thus permit successively deeper layers of subdivisions, revealed as the user sees fit.

a) First level of display (upon selection of main term)

Browsing subject term: ISLAM
n citations found under the term ISLAM
Term is also subdivided by the following categories:

1. Congresses
2. Periodicals
3. Subdivided by geographical area, e.g., Afghanistan
.
.
[user selects 3]

b) Second level of display:

Browsing term: ISLAM--SUBDIVIDED BY GEOGRAPHICAL AREA
n citations found under the term ISLAM -- SUBDIVIDED BY GEOGRAPHICAL AREA
Term is subdivided by the following geographical areas:

1. Africa
2. Africa, East
3. Africa, North
4. Africa, West
.
.
[user selects 2]

c) Third level of display:

Browsing term: ISLAM--AFRICA, EAST
N citations found under the term ISLAM--AFRICA, EAST
Term is also subdivided by the following categories:

1. Bibliography
2. Congresses
3. History

Fig.4 Possible Levels of Display

2. Specify a matching routine that will link a main heading that has been subdivided by both direct and indirect subdivision practice. Since the practice of direct geographic subdivision ("Fountains--Rome") has been replaced in recent years with indirect geographic subdivision ("Fountains--Italy--Rome"), these headings would currently appear dispersed in an alphabetical display. Therefore, when two *z* subfields are present (indirect subdivision), the character string in the second *z* subfield can be matched against an identical occurrence appearing in the first *z* subfield (direct subdivision). This would cluster all identical geographic areas together regardless of past practice.

3. Since most form subdivisions are made redundant by coding that exists in MARC fixed fields, the system might simply ask if a list of formats used with the topic is desired, rather than displaying the formats routinely. Alternatively, a table could be specified to cluster subdivisions that represent the type or format of the work.

4. Consider abandoning certain free-floating subdivisions altogether, such as LC has already done with "Addresses, essays, lectures" and "Yearbooks". Subdivisions of this type add nothing to increase a subject's specificity and considerably hamper effective subject retrieval by arbitrarily separating records that could otherwise be clustered together.[7]

5. Tables would be operative if a user selects a subject term or phrase when the number of index entries exceeds a predetermined limit; otherwise, there is no need to compress entries. Non-free-floating subdivisions could be treated as "specialized aspects" for purposes of limiting the display after a search item has been specified. Researchers must closely examine the specific topical subdivisions to see if any significant redundancy of concepts emerges that would be suitable for compression and to identify areas that will prove incompatible with such an approach.

These are only some of the characteristics of subdivisions that need to be investigated closely if we aim to gain control of the data existing in MARC records and to manipulate it with the user in mind. Imaginative solutions to problems may begin to surface once we are able to reduce the noise levels that are present in current browsable displays.

Where we can go from here

In designing effective browsable subject index displays, we must move toward examining compression techniques to limit the size of displays presented to the user. The use of tables must also be examined as an effective method of handling the existing redundancy of concepts present in the LC subdivisions, substituting textual messages for groups of conceptually similar terms.

On a larger scale, an expanded MARC record could greatly aid the manipulation of data in subject heading fields. The MARC record to date has not been exploited with regard to subject access, although improvements have been suggested.[8] The *x* subfield code should be replaced by a detailed, mnemonic coding scheme designed to cluster similar *concepts* together (for example, subfield *m* = subdivisions for mathematical aspects). Ironically, the specificity that subdivision has tried to provide has been obscured in the MARC record by a general subfield code that cannot be manipulated effectively in the machine environment.

---

[7] An interesting, but curious case in point is a monograph cited in this paper, *Redesign of Catalogs and Indexes for Improved Online Subject Access* by Pauline Cochrane. LC has assigned an unprecedented six subject headings, all subdivided by "Addresses, essays, lectures," but this subdivision does nothing to increase the subjects' specificity. Ironically, LC has missed the point entirely by failing to assign to this seminal collection of papers the specific subject heading "Catalogs, On-line--Subject access" (which first appeared in the 1982 supplement to LCSH).

[8] Cochrane, *Improving LCSH for Use in On-line Catalogs*, p.62-68; Pauline Atherton, *BOOKS are for use: final report of the Subject Access Project to the Council on Library Resources* (Syracuse, N.Y.: Syracuse Univ. School of Information Studies, 1978); Pauline A. Cochrane and Karen Markey, "Preparing for the Use of Classification in Online Cataloging Systems and in Online Catalogs," *Information Technology and Libraries* 4:91-111 (June 1985); Carol A. Mandel and Judith Herschman, "Online Subject Access: Enhancing the Library Catalog," *Journal of Academic Librarianship* 9:148-55 (July 1983); Mandel, "Enriching the Library Catalog Record for Subject Access," in *Improving LCSH for Use in Online Catalogs*, p.231-40; and Arnold S. Wajenberg, "MARC Coding of DDC for Subject Retrieval," *Information Technology and Libraries* 2:246-321 (Sept.1983).

The technological achievements possible with online systems have great scope for subject access. The browsable display is one area that will respond favorably to our immediate attention, and libraries do not need to wait for action on a national scale to proceed with improvements. Computes have sorting capabilities that can be exploited more creatively than simply routinely listing every subdivision from A to Z. Clustering conceptually similar terms can reduce the size of a large display while simultaneously providing a structure that will assist a subject search. The architecture of online public access catalogs must be designed to place the burden of data organization on the technological capabilities of the system (which is exactly the type of thing computers are, after all, designed to do), rather than resting it with the user.

We must attempt to design a conceptual framework that will assist a user quickly and intuitively, something that was never really possible in a card catalog due to its physical and logistical limitations. It is incumbent upon the profession to speedily discover new ways of manipulating our existing encoded MARC records to rectify some of the problems inherent in our current online systems. Computing power allows us to invent innovative solutions to problems that the profession could not adequately solve before, but this cannot be done unless we unload some of our preconceptions of how catalogs are supposed to be constructed and what users expect of a library catalog. We must step back and approach our dual tasks of bibliographic control and information retrieval with insights made possible by state-of-the-art technology.