INDEXING A CLASSIFICATION SCHEME

J. MILLS

The store of information which constitutes a library differs from the store of information in one book in some obvious respects. It is much more extensive in content, and it consists physically of thousands—perhaps hundreds of thousands—of discrete items, any one of which might be sought through numerous different approaches; in particular, it may be sought through its author, its title, or its subject. Any one book, moreover, as indexers well know, may bear on numerous different subjects or themes. Ideally, if a library is to function perfectly, every item of information in its possession should be retrievable on demand, rapidly and with certainty. The function of the library is basically the storage and retrieval of information.

The problem of mere numbers raises, as always, problems of organization, and virtually all libraries at some point or another turn to systematic arrangement, i.e. classification, either of the physical stock itself, or of the catalogues to it, or of both. Here, for example, is a small part of Dewey's Decimal Classification, which is designed to serve such an arrangement and which is the most widely used scheme in the world and the basis of the UDC (many classes have been left out, of course; this example seeks only to show certain features of relative sequence of classes):

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>Art</td>
</tr>
<tr>
<td>704.94</td>
<td>Subjects in art (Iconography)</td>
</tr>
<tr>
<td>704.948</td>
<td>Religious</td>
</tr>
<tr>
<td>709.02</td>
<td>Medieval</td>
</tr>
<tr>
<td>709.44</td>
<td>French</td>
</tr>
<tr>
<td>709.4401</td>
<td>Medieval</td>
</tr>
<tr>
<td>720</td>
<td>Architecture</td>
</tr>
<tr>
<td>720.944</td>
<td>French</td>
</tr>
<tr>
<td>726</td>
<td>Religious</td>
</tr>
<tr>
<td>750</td>
<td>Painting</td>
</tr>
<tr>
<td>751.42</td>
<td>Water-colour</td>
</tr>
<tr>
<td>755</td>
<td>Religious</td>
</tr>
<tr>
<td>755.0944</td>
<td>French</td>
</tr>
<tr>
<td>755.094401</td>
<td>Medieval</td>
</tr>
<tr>
<td>758.1</td>
<td>Landscape</td>
</tr>
<tr>
<td>758.1[1]</td>
<td>Romantic</td>
</tr>
<tr>
<td>758.10942</td>
<td>English</td>
</tr>
<tr>
<td>758.10942[1]</td>
<td>Water-colour</td>
</tr>
<tr>
<td>759.02</td>
<td>Medieval</td>
</tr>
<tr>
<td>759.4</td>
<td>French</td>
</tr>
<tr>
<td>759.401</td>
<td>Medieval</td>
</tr>
<tr>
<td>759.4C6</td>
<td>Corot</td>
</tr>
<tr>
<td>759.4C6[1]</td>
<td>Landscapes</td>
</tr>
</tbody>
</table>
The purpose of this sort of arrangement is fairly clear; it brings together in a helpful way material which is closely related. The books on Painting, say, are followed by those on particular media of painting (e.g., Water-colour), on subjects of painting (e.g., Religion), on periods of painting (e.g., Medieval), and on national schools of painting (e.g., French), and on individual painters (e.g., Corot). But it is also clear from the example that there are definite limits to the degree to which related topics are brought together; e.g., the reader wanting material on Medieval art (its painting, its architecture, its national schools, etc.) is not going to find it all together. Similarly, material on French art, or Religious art, will be scattered, not collected.

A second problem raised by classification is that its order, whilst very helpful, is not particularly obvious; e.g., how is the reader to know that French Landscape Painting is to be sought under Landscape Painting, not under French Painting? If the sequence of books in a general library is considered, the problem assumes alarming proportions. A run of 50,000 books occupies something like half-a-mile of shelving. Where, in such a sequence, would the subject of, say, * Strikes in the Coal-mining Industry of Great Britain* be found? Economic history of Britain? Or Economics of the Coal Industry? Or under Coal-mining in Technology, with a sub-class "Economic History"? Or under Strikes, in Labour economics? And where does economics, say, come in a sequence of fifty other major classes? (Indexers may reflect that this is no worse than the problem of finding just where, in a 500-page treatise, a particular person or place or subject is mentioned.)

This problem is solved by two devices. Firstly, by attaching a notation—i.e., numbering every term (representing some class or sub-class) so that its relative position is mechanically established; e.g., 331.89282233 is the rather forbidding number for Strikes in the Coal-mining Industry in the Decimal Classification. Secondly, by providing an A/Z Index whereby the reader can look up the name of the subject required and be given its class number, e.g.—

- Coal-mining: Economics 338.272
- Coal-mining: Engineering 622.33
- Coal-mining: Strikes: Economics 331.89282233

This example shows that such an alphabetical index, besides providing a rapid key to the location of topics, also goes a long way towards solving the first problem (that a classification inevitably scatters some related topics). For it automatically shows those different aspects of a topic which are not brought together in the classified sequence. In fact, the A/Z Index constitutes a second classification of concepts in which, under the name of concrete subjects like Coal-mining, are listed all the varied contexts in which it may be found. The modern view of a classification system in the library is that it is two interlocking classifications; and the alphabetical one is an integral part of the system, not just an appendage.

(It may be noted here that although indexing in its widest sense covers all indicating, or guiding, of literature, it is used here in the restricted sense of being an alphabetical complement to the classified order. Also, a library provides alpha-
betical indexes to the authors (including editors, illustrators, etc.) and perhaps to
the titles of books, as well as to the subjects. But here, we are concerned only
with an index to the thought-content of the books.)

Making an index

The problems of indexing a classification flow from the fact that readers con-
sulting the relatively complicated hierarchy which constitutes a classification may
approach it from many different directions; e.g., a reader looking for material on
French Landscape Painting may look in the index under French Landscape Paint-
ing, or French Painting, or Painting, or Landscape Painting, or even Art, and so on.
(This is in addition to the familiar problem of simple synonyms, such as Ecclesiasti-
cal Architecture and Religious Architecture.)

One way of ensuring that all these approaches are indexed would be to prac-
tice elaborate permutation of the several terms which make up the name of all such
compound topics (which are very numerous in literature), e.g., French Landscape
Painting; Landscape painting, French; Painting, French, Landscape and so on.
This is a daunting prospect. If such a policy were pursued at all consistently it
would produce very large and unwieldy indexes. But it is most unlikely that con-
sistency would be maintained, for some indexers would inevitably tend to permute
less generously than others.

The method known as chain indexing solves this problem by ruling that in a
compound entry (one consisting of two or more terms) the terms must be cited in
a single, definite order only. This will be shown in the examples below. A chain,
in library classification, is a statement of the successive steps of division whereby a
class is broken down into successive sub-classes, sub-sub-classes, etc.; e.g., in the
Decimal Classification, Art is divided into different media (Architecture, Painting,
etc.) then Painting (say) is divided into various subjects (Portraits, Landscapes,
etc.) then Landscape painting (say) is divided into various schools (French, Eng-
lish, etc.) The subject French landscape painting, therefore, has the chain Art—
Painting—Landscape—French. Each of these “levels” of the classification reflects
a possible approach to the subject, so each one is indexed in turn. For example, if
a book on French Landscape Painting were the first book to be indexed in the
library it would be first analysed as a chain, and then the term representing each
step would be used as the first word in an index entry:

700 Art would give the entry Art 700
750 Painting would give the entry Painting 750
758.1 Landscape would give the entry Landscape painting 758.1
758.10944 French would give the entry French landscape painting 758.10944

It can be seen that the indexing term (the term appearing first in an index
entry) may need to be qualified by other terms in the chain in order to show clearly
the particular context involved.
e.g., to distinguish Landscape painting 758.1
Landscape gardening 712
Landscape drawing 743.82
Landscape: Physical geography 551.4

It is a fundamental rule of chain indexing that qualifying terms must be taken from terms above the indexing term in the chain, and never from terms below. The terms below represent divisions of the subject (e.g., in the chain above, Landscape is a division of Painting, not vice-versa; French Landscape Painting is a division of Landscape Painting, but Landscape Painting is not a division of French Painting). If qualification by less general terms were tolerated, then these would be the sort of entries resulting for the above field:

- Landscape painting, English 758.10942
- Landscape painting, French 758.10944
- Landscape painting, German 758.10943
- Painting, Landscape 758.1
- Painting, Portrait 757 [1]
- Painting, Religious 755

It can be seen that what is happening is simply a repetition of what the classified order already does; i.e., the Decimal Classification divides Landscape Painting (758.1) into its various schools (758.10941, 758.10942, 758.10943, etc.), or Painting into its various subjects (753, 754, 755, etc.) So a reader who really wants a book on French landscape painting, but looks in the Index under Landscape, would be quite adequately served by the reference Landscape painting 758.1 and nothing else. For as soon as he consults this number on the shelves or in the classified part of the catalogue (where the full entries for the individual books appear) he will find the further divisions he is seeking. This assumes, of course, that both shelves and catalogue are well guided. Similarly, a reader wanting a book on Landscape Painting, but looking under the term Painting, will be adequately served by the simple entry Painting 750. The absence of an entry Painting, Landscape 758.1 alongside this will not prevent him finding this division of 750 when he gets to that number on the shelves or in the classified catalogue.

It should not be thought that every book will need so many entries as the first one demonstrated above. If the second book to be indexed in our hypothetical library were on English landscape painting, the only index entry necessary for it would be English landscape painting 758.10942; for the rest of the chain would already have been indexed. In fact, a situation is soon reached when no new index entries at all are necessary for many books.

It may be objected that the library may not possess a book on Landscape Painting in general, yet carry the index entry Landscape Painting 758.1. But this is not a misleading entry, in that if the library possesses a book on French landscape painting, or English landscape painting, etc. (which it must have, to have
given rise to the entry) this is material on Landscape painting—perhaps the very part of the subject the reader really had in mind. It should be realized that the meaning of any index entry is that something on the topic indexed will be found at, or following, the number given.

Chain indexing, then, is a very powerful aid in achieving the major aims of the indexer of a classified sequence, which are: firstly, to make sure that every keyword (i.e. every term likely to be consulted by the readers of the library concerned) is indexed (i.e., appears as the first word of some entry); secondly, because of the numerous ways in which many subjects can be indexed, to do this systematically, so as to avoid doubt and inconsistency on the part of the indexer; thirdly, to do this economically, avoiding superfluous entries by recognizing that many of the pathways to the location of information are provided by the classified sequence itself, and need not be repeated in the alphabetical index.

To demonstrate further how this is done, and to clarify various points of practice, here are the index entries which would be made to index the hierarchy given earlier:

<table>
<thead>
<tr>
<th>Entry</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture</td>
<td>720</td>
</tr>
<tr>
<td>Art</td>
<td>700</td>
</tr>
<tr>
<td>Corot, Camille: Painting</td>
<td>759.4C6</td>
</tr>
<tr>
<td>Ecclesiastical architecture</td>
<td>726</td>
</tr>
<tr>
<td>English landscape painting</td>
<td>758.10942</td>
</tr>
<tr>
<td>Fine arts</td>
<td>700</td>
</tr>
<tr>
<td>France: Architecture</td>
<td>720.944</td>
</tr>
<tr>
<td>French art</td>
<td>709.44</td>
</tr>
<tr>
<td>French painting</td>
<td>759.4</td>
</tr>
<tr>
<td>French religious painting</td>
<td>755.0944</td>
</tr>
<tr>
<td>Iconography</td>
<td>704.94</td>
</tr>
<tr>
<td>Landscape painting</td>
<td>758.1</td>
</tr>
<tr>
<td>Landscape painting: Corot</td>
<td>759.4C6[1]</td>
</tr>
<tr>
<td>Medieval art</td>
<td>709.02</td>
</tr>
<tr>
<td>Medieval French art</td>
<td>709.4401</td>
</tr>
<tr>
<td>Medieval French painting</td>
<td>759.401</td>
</tr>
<tr>
<td>Medieval French religious painting</td>
<td>755.094401</td>
</tr>
<tr>
<td>Painting</td>
<td>750</td>
</tr>
<tr>
<td>Religious architecture</td>
<td>726</td>
</tr>
<tr>
<td>Religious art</td>
<td>709.948</td>
</tr>
<tr>
<td>Religious painting</td>
<td>755</td>
</tr>
<tr>
<td>Romanticism: Landscape painting</td>
<td>758.1[1]</td>
</tr>
<tr>
<td>Water-colours: English landscape painting</td>
<td>728.10942[1]</td>
</tr>
<tr>
<td>Water-colours: Painting</td>
<td>751.42</td>
</tr>
</tbody>
</table>
Some further points of interest are:

(i) Synonyms are indexed directly to the number, to save the reader further searching.

(ii) A standardized style (nouns in the plural, separated by colons) is followed where possible; e.g., to have numerous entries beginning with Hospital (using it adjectivally), followed by numerous entries, beginning with Hospitals, overlapping those in front, and the two sets separated, perhaps, by terms like Hospitalers: Orders of Knighthood and Hospitality: Ethics is avoided by this practice. But it should be noted that subjects in Art are not typical here, in that adjectival forms are usually preferred, as above.

(iii) Steps of division represented by words which are not likely to be sought by readers (e.g., Theory, Organization) are omitted, or are given a general reference (e.g., Dictionaries on special subjects, see subject, instead of, say, Dictionaries: Chemistry 540.3). Period divisions usually come into this category of “unsought links” in the chain; but “Medieval” is often regarded as an exception. This example emphasizes that chain indexing does not dispense with the individual discretion of the indexer, or the need for him to be alive to the terms used by the readers.

(iv) The analysis of the subject which must precede the making of index entries for it (but which is only rough work, not seen in the finished index, of course) in no way depends on the notation (i.e. the class-numbers), e.g., in the Decimal Classification, one chain is:

5  Science
574  Biology
59  Zoology
596  Vertebrates
599  Mammals

It can be seen that the numbers do not express the relations between the subjects represented (e.g., 59 does not appear to be a division of 574). Again, 704.94 Iconography looks as though it is a division of 704.9, which actually is Collections of Essays on Art. Similarly, if the classification system fails to give a precise number for a topic, this does not mean that it will go unindexed, e.g., there is no number for Romanticism under Landscape Painting in the Decimal Classification; but the index entry has been made above to the nearest number. The [1] is added to show that it is a division of the subject, not the general subject represented by the number alone.

(v) Inconsistencies in the classification system do not nullify the value of the index, e.g., the Decimal Classification locates Subjects of Art ahead of National Schools of Art—but under Architecture this order is reversed, for no particular reason; this will not stop the index entries guiding the reader to the correct place, however.
(vi) The physical form of the index is a matter of choice. Ideally it should be infinitely flexible—i.e., allow for constant additions and withdrawals. This suggests cards as the ideal form. Typewritten sheets in loose-leaf folders are easier to consult, but additions are less easily made and re-typing of individual sheets is inevitably necessary from time to time.

(vii) Many libraries do not bother to make an alphabetical index to their own stock and use instead the printed index accompanying the scheme itself. There is growing recognition of the fact that this gives a very poor indexing service, for these indexes are primarily tools for the cataloguer, not the reader, and are always out-of-date, completely inaccurate as a guide to subjects represented in the library (indexing thousands of topics not in the library and omitting thousands which are), as well as using (usually) American terms rather than English ones.

Chain indexing and the indexing of books

Many books display a definite plan in their presentation of information, and some text-books display in their contents table a virtual classification of the subject and its divisions. On the other hand, the information in a book is rarely presented directly in the form of a detailed classification of the subject covered and so the relevance of chain procedure to the indexing of individual books does not appear to be very marked. But the central principle is undoubtedly relevant when deciding the degree and manner of permutating the terms of compound headings (e.g., Electronic Indexing of the Dead Sea Scrolls). This principle is that the alphabetical index need not attempt to duplicate the grouping and association of one topic with another which is found in the text of the book itself. This point may be demonstrated by the index of a text-book on library classification which the writer prepared recently.

The plan of the book was to deal first with the theoretical principles of library classification, including such topics as order of classes, subject analysis, notation (its brevity, flexibility, etc.), alphabetical indexing, etc. Then six major classification schemes were described and the problems just mentioned were considered again as they occurred in each scheme in turn. In effect, this meant that nearly everything on a given scheme was considered together under that scheme, but information on a particular problem (e.g., Notation) was scattered; the chapter on Notation dealt with general principles and features—but not with the notation of individual schemes (except insofar as examples of particular notational problems drew on these for demonstration purposes). This situation was analogous to a simple classification pattern:

Library classification
(Problems)—e.g., Subject analysis
  Notation
  A/Z index
(Schemes)—e.g., Decimal Classification
Subject analysis
Notation
A/Z index

Universal Decimal Classification (UDC)
Subject analysis
Notation
A/Z index
etc.

This suggested that in the index to the book, the index entry for each scheme would need to be only a single reference to the chapter on that scheme, with the implication that everything on it would be found following the number given. On the other hand, the index entry for a particular classification problem would need to refer to all the numerous places in which the problem was dealt with—i.e., a reference to the general chapter in the first half of the book, and references to it as it appeared in the several schemes, e.g.:

BC see Bliss, H. E. Bibliographic Classification
Bibliographic Classification see Bliss, H. E. Bibliographic Classification
Bliss, H. E.
  Bibliographic Classification, 133ff., 162
  brevity in notation, 39-41, 172
    in BC, 135, 144; CC, 125; DC, 63-4; LC, 96-7; SC, 110-1; UDC, 79-80
    [the abbreviations BC, CC, DC, etc., stand for the full names of the schemes, e.g., Bibliographic Classification, Colon Classification, Decimal Classification.]
  expressive notation, 38, 39, 40-1
    in BC, 144; CC, 124, 125; DC, 63, 68; LC, 97; SC, 100; UDC, 79.
  notation, vii, 16, 32, 37ff., 174, 178
    affecting order, 22, 38-9, 47
      in BC, 138-9; CC, 124; DC, 63; LC, 96; SC, 105; UDC, 79

It should be noted that under the entry “Bliss . . .” there is no attempt made to repeat the references to that scheme found under the individual problems (e.g., “brevity in notation . . . in BC, 135, 144 . . .”; or, “expressive notation . . . in BC, 144 . . .”). The one reference, to 133ff., will locate the main chapter on the scheme, in which all these problems will be found dealt with. Similarly, but at a less general level, the entry: “Notation . . . 37ff. . . .” will locate the main chapter on that problem, in which are considered the various features of notation, such as brevity, expressiveness. So there is no need to duplicate, by inversion, the entry for “brevity in notation”, or for “expressive notation”—i.e., no entry is made for “notation, brevity in” or for “notation, expressive”. 

47
It may be remarked that the bare entry “Bliss, H. E. Bibliographic Classification, 133ff.” does not give as rapid access to the location of a particular problem (e.g., Brevity of Notation in the BC) as the direct reference “Brevity . . . in BC, 135, 144”. This may be compensated for in two ways. Firstly, the contents table (which may be regarded as being complementary to the alphabetical index) might display an analysis of each chapter. Or, secondly, modification of the chain indexing principle might be made to produce such an outline analysis in the index, e.g.:

Bliss, H. E.
Bibliographic Classification, 133ff., 162
facet analysis, 142-4
index, 147 [i.e., the printed index]
indexing, 149ff. [i.e., making an index to a collection arranged by BC]
notation, 144ff.
order and collocation, 135ff.
practical work in, 147ff.
principles, 133ff.

This last treatment was in fact the one followed in the book in question. But the writer is not sure that a detailed and analytical Contents statement at the beginning might not be better.

INDEXING AS AN ART

An article by “Strix” entitled “The Last Lap” and published in The Spectator contains an extract from How to be useful and happy from sixty to ninety, by A. Lapthorn Smith (Bodley Head), 1922. “Strix” writes: “. . . its index is strongly to be recommended and may indeed be said to stand in a class by itself.” His sample follows:

Alcohol as cure for insomnia, very bad 190
Beard, long white, don’t wear 56
Berens, Mrs., still wins prizes at archery at 77 30
Bulky food exercises bowels 133
Carriage and pair shortens life 65
Cook, good, source of danger to elderly men 69
Crime to die rich 112
Engine drivers over sixty, what to do with 44
If no relatives, spend on poor 45
M.M., Miss, of Kent, just 70, feels 35 60
Otherdom leads to happiness 48, 53
Rich man, difficult for, to reach 100 67
Young people, company of, at 60, how to keep 56

(Cutting contributed by A. B. Lyons)