THE USE OF CLASSIFICATION IN BOOK INDEXING

D. W. L. LANGRIDGE *

It is my contention that no indexer can do his work unless he has in mind at least an implicit classification of the subject he is dealing with. If this is so, then he will obviously lose nothing and probably gain a lot by making this classification explicit. The following article is based on experience of following up this idea in the field of social science.

Keys to knowledge

The objective of the indexer, whether he is dealing with a single book or with a collection, is to provide a key to the knowledge contained therein. It is essentially an educational objective, for an effective guide must show not only the location of specific items but also the relationships obtaining between them. For example, as well as including the specific items “Controlling”, “Managing”, “Organizing” and “Planning” it should show that they are related as follows:

\[
\begin{array}{c}
\text{Management} \\
\text{Planning} & \text{Organizing} & \text{Controlling}
\end{array}
\]

In a library this end is most satisfactorily achieved by arranging the books in a classified order which shows the chief relationships and by providing an alphabetical index as a key to the location of specific items. The book also has an alphabetical index and the text corresponds to the collection of books in the library. This text will be at the best in a very roughly classified order and, as we are dealing here with very minute items of knowledge in comparison with those in the library, we shall find a great deal of scatter of identical and related subjects throughout the text. Since, then, the order of items in the text of a book is nowhere near as systematic as that of books in a library, the burden of showing relationships must be taken by the index if it is to provide as effective a key to the knowledge in a book as the combined classification and index do for a library.

Relationships between terms

Most book indexes contain some cross-references showing relationships, but it is difficult to see how this work can be carried out completely systematically if the normal procedure is followed of filing entries alphabetically during compilation of the index. The alternative I am suggesting is to file entries in a classified order during compilation, leaving alphabetical ordering until the very end, when all useful relationships have been noted and suitable entries made.

What are the relationships we are likely to be concerned with? We may divide them into two kinds: the basic relationships between simple subjects and the relationships between the various parts of a compound subject.

The basic relations are (1) generic to specific and whole to part, (2) specific to generic and part to whole, (3) co-ordinate. To illustrate this with a simple

* Based on paper read to the Society in November, 1960.
example, in a book containing information on Africa as a whole as well as information on individual countries in Africa we should need the following types of reference:

1. Africa, see also Ghana, etc. (whole to part)
2. Ghana, etc., see also Africa (part to whole), and possibly
3. Rhodesia, see also Nyasaland (co-ordinate)

That there is some need to make the provision of this type of reference more systematic is suggested by a chapter in the second edition of R. L. Colli-son's Indexes and Indexing. Here he calls in Boolean algebra to assist in showing relationships, producing a diagram on page 71 in which A might be interpreted as Locomotives, B as Railways, and C as Station restaurants. But surely this subject is seen more clearly if analysed into its different aspects in the way that it would be in a classification scheme, e.g.:

RAILWAYS

<table>
<thead>
<tr>
<th>Rolling stock</th>
<th>Fixed equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locomotives</td>
<td>Signal boxes</td>
</tr>
<tr>
<td>Steam</td>
<td>Stations</td>
</tr>
<tr>
<td>Electric</td>
<td>Platform</td>
</tr>
<tr>
<td>Diesel</td>
<td>Waiting room</td>
</tr>
<tr>
<td>Wagons</td>
<td>Restaurant</td>
</tr>
<tr>
<td>etc.</td>
<td>etc.</td>
</tr>
</tbody>
</table>

It is quite obvious here that, while references would be required in both directions between locomotives and rolling stock, there would not be any basic relation between locomotives and station restaurants.

A further example from this chapter of Indexes and Indexing, page 70, is not only less clear than a classification would be but is also quite misleading as to the nature of the relations between parts. If A is interpreted as Theatre, B as Television, and C as Stage representation of drama, the common area X would represent Televising of stage representation of drama. If one made a classification scheme to cover the book from which this particular topic was taken it might include somewhere the two following groups of subjects:

<table>
<thead>
<tr>
<th>Literature</th>
<th>Communication media</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forms</td>
<td>Primary</td>
</tr>
<tr>
<td>Drama</td>
<td>Theatre (stage)</td>
</tr>
<tr>
<td>etc.</td>
<td>Cinema</td>
</tr>
<tr>
<td></td>
<td>etc.</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
</tr>
<tr>
<td></td>
<td>Radio</td>
</tr>
<tr>
<td></td>
<td>Television</td>
</tr>
<tr>
<td></td>
<td>etc.</td>
</tr>
</tbody>
</table>

This kind of analysis surely makes much clearer the relationship between the various terms.

**Compound subjects**

The problem of compound subjects may be illustrated by the following analysis of education and training:
Many of the compound subjects occurring in this field will contain only two elements from this list, but a combination of more than two is possible, in fact the whole five is possible as in the following example:

Training/middle managers/for top management/by teaching them economics/at University/summer schools/in England.

To write down the categories occurring in a subject in their order of importance can be a great help in dealing with a complex field such as this one. Moreover it brings to light concealed elements. For example, it can be seen from the table above that the term management could occur in three different categories, viz., the people taught, the purpose of teaching, or the subject matter. Thus the phrase management training could mean:

1. Training managers (i.e., for further stages in their career)
2. Training people who are not already managers to be managers
3. Training people in the subject matter of management
or even a combination of these such as: "training managers for higher stages of management by teaching them the subject matter of management". We cannot index correctly unless we do know which of these meanings the term bears at any one time and it would be difficult to find out without this sort of analysis to aid us.

Another possible concealed element is the country. A book written in Great Britain for example and dealing with other countries as well may not specifically mention Great Britain every time it applies to a statement, but these implicit uses should be noted while indexing. Of course in all these matters the indexer must finally use his judgment as to what is necessary in the index. The important thing is that if any entries or parts of entries are to be left out it must be done consciously and not from failure to recognize their existence. As long as a fixed order of parts in compound subjects is observed the indexer need make only one entry during compilation of the index, leaving decisions about desirable permutations until the end.

Another advantage of the analysis of categories in a subject is that it shows quite clearly the functions of two groups of vague terms. The first is composed of words such as "effect" and "influence" which merely express relationships between essential terms and therefore do not require entries themselves; the second contains words like "methods" and "types" which are the names of categories in the subject. It is doubtful whether readers normally seek such general terms, but there is no reason why one should not have an introduction to the index such as the following from a book on direct mail selling:
“This index contains reference to
(a) all direct mail techniques referred to in the text
(b) all goods and services used in illustration of direct mail techniques
(c) all markets approached
(d) all firms and persons mentioned.”

Procedure

One may first distinguish between those entries that can be taken directly from the text and those which have to be chosen to represent the content of a passage. The first will include such items as the names of people and organizations and the titles of books. They are comparatively easy entries to deal with and may be treated quite separately—during the preliminary run through the book if desired. To revert to the parallel with library procedure, this group corresponds to the author or name catalogue. The remainder are our subject entries which require classification. The alternatives are either to use a suitable existing classification or to draft the outline of a scheme that will cover the subject matter of the book. The contents list and introductory chapter should be a great help in deciding what categories are manifested in the subject and once they are decided it is easy to see what kind of terms will occur in each.

By this method it should frequently be possible to prepare for the compilation of the index without the usual detailed reading that is recommended; and since time is still the greatest enemy of the indexer this is a definite advantage of the method. It may also be possible to reduce the time of compilation by dividing the different types of entry between more than one worker.

It is worth mentioning in conclusion that if one ever has to deal with an unfamiliar subject the technique of analysing into categories is invaluable for obtaining a quick understanding of its structure.

INDEXERS ALSO?

It is the fate of those who toil at the lowest employments of life, to be rather driven by the fear of evil, than attracted by the prospect of good; to be exposed to censure, without hope of praise; to be disgraced by miscarriage, or punished for neglect, where success would have been without applause, and diligence without reward.

Among these unhappy mortals is the writer of dictionaries; whom mankind has considered, not as the pupil, but the slave of science, the pioneer of literature, doomed only to remove rubbish and clear obstructions from the paths through which Learning and Genius press forward to conquest and glory, without bestowing a smile on the humble drudge that facilitates their progress.

Every other author may aspire to praise; the lexicographer can only hope to escape reproach, and even this negative recompense has yet been granted to very few.

From Preface to Dr. Samuel Johnson’s Dictionary.

(Contributed by L. E. C. Hughes.)