Cross-references in back-of-book Indexes

Virgil Diodato

An analysis of 447 books in arts/humanities, science/technology, the social sciences, and generalities determined how frequently and for what purpose cross-references were used in back-of-book indexes. Of 376 indexes examined, 300 included cross-references. The average index had 32 cross-references per index or 96 cross-references per 1,000 text pages. Of the three major subject areas, cross-references most commonly occurred in science/technology indexes. 'See' references were surprisingly dominant over 'see also', especially in science/technology and the social sciences. The references made many kinds of links, most often to show synonymy, class membership, and agent/activity relationships. It is encouraging that many indexes have cross-references, but name indexes could have more.

Introduction

Why do back-of-book indexes use cross-references? How often do they use this powerful tool?

Cross-references are part of the syndetic system of an index. Syndetics can also include glosses, inversions, and other devices that explain the meanings of index terms or help a user find the most appropriate index terms. One also may think of multiple entries as a syndetic device or at least a substitute for cross-references.

We do not have to be convinced of the importance of syndetics. Milstead tells us that the syndetic structure in an index is 'an extremely useful adjunct to index use'. Borko and Bernier note that '[a]n index without a syndetic system is inferior to an index with such a system'. According to the American National Standards Institute's Basic criteria for indexes, cross-references and other syndetic structures tie 'the various parts of the index together into a unified whole'. Even in hypertext documents, Brown assures us that the system 'should support cross-reference links'.

Cross-references have at least two major functions. First, they send readers to the preferred form of an entry, as in 'Doctors: see Physicians'. Second, they suggest additional indexing terms to the reader, as in 'Physicians: see also Hospitals; Medical schools'. In both cases the cross-reference makes a link between the term that the reader looked up and some other term in the index.

This study

I studied 376 recent back-of-book indexes to find out how often back-of-book indexes used cross-references and what kinds of links the references established. The books came from a variety of subject fields and were indexed probably by a variety of people, including not only professional indexers but also authors, editors, and others.

Method

An indexing student (Ms Georgianna Henry) and I randomly sampled one book title from each of 1,126 pages of the 1988 cumulation of the American book publishing record (excluding the 'Fiction' and 'Juvenile Fiction' sections at the rear). We looked for the books in the catalogs and on the shelves of three academic libraries in Milwaukee, WI: the Marquette University Memorial Library, the Marquette University Science Library, and the University of Wisconsin–Milwaukee Golda Meir Library. We found 502 of the books in the catalogs and actually were able to examine 447 of these. Three hundred and seventy-six (84%) had indexes. I recorded the size of the indexes, the number of cross-references, and the nature of the links made by them.

Previous studies

Several earlier studies collected information about the characteristics of back-of-book indexes, and in the process collected data about their cross-references. Gratch, Settel, and Atherton analyzed 113 back-of-book indexes in the humanities and social sciences and found that 67 (59%) had cross-references. In a small study in various subject fields, Gandt and I found that of 37 books indexed by their authors, 30 (81%) had cross-references, while all 27 (100%) books indexed by others, many of whom apparently were professional indexers, had cross-references. Nwodo and Otokunefor found that the rate of including cross-references in Nigerian books decreased in the 1980s compared to the 1960s. Bishop and Liddy found cross-references in 80% of 433 back-of-book indexes in various subject fields.
CROSS-REFERENCES IN BACK-OF-BOOK INDEXES

Table 1. Number and size of indexes

<table>
<thead>
<tr>
<th>Subject categories and LC classes</th>
<th>Arts/humanities</th>
<th>Science/technology</th>
<th>Social sciences</th>
<th>General/bibliog.</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B, P, M, N</td>
<td>Q-V</td>
<td>C-L</td>
<td>A, Z</td>
<td></td>
</tr>
<tr>
<td>No. of books in sample examined</td>
<td>279</td>
<td>400</td>
<td>421</td>
<td>26</td>
<td>1,126</td>
</tr>
<tr>
<td>with indexes</td>
<td>84</td>
<td>133</td>
<td>184</td>
<td>12</td>
<td>447</td>
</tr>
<tr>
<td>% with indexes</td>
<td>71</td>
<td>91</td>
<td>88</td>
<td>75</td>
<td>84</td>
</tr>
<tr>
<td>Mean pages of index*</td>
<td>8.7</td>
<td>9.9</td>
<td>9.0</td>
<td>24.3</td>
<td>9.6</td>
</tr>
<tr>
<td>Mean pages of text*</td>
<td>293.3</td>
<td>373.0</td>
<td>287.1</td>
<td>279.3</td>
<td>315.9</td>
</tr>
<tr>
<td>Mean number of index pages per 1,000 text pages</td>
<td>28.7</td>
<td>27.5</td>
<td>31.5</td>
<td>99.0</td>
<td>31.2</td>
</tr>
</tbody>
</table>

*Calculated only for the 376 books with indexes

into the three major subject fields (arts/humanities, science/technology, and social sciences) and a general/bibliography area. There was a much lower proportion of books with indexes in arts/humanities and general/bibliography than in the other two fields; only nine indexed books in the general/bibliography area, all in class Z.

How large were the indexes?

I compared the number of pages in all the indexes in each book to the number of pages in its text. Of the three major subject fields, this ratio was largest for the social sciences: 31.5 index pages per 1,000 pages of text. Books in science/technology had the smallest ratio, 27.5. See Table 1.

What types of indexes were analyzed?

Most of the books (334 of 376) had only one index. If there was more than one type of index, I analyzed only the subject (subjects only) or general (subjects and names) index. The analysis included 344 subject or general indexes and 32 name indexes. See Table 2.

The 300 indexes containing cross-references

Eighty per cent (300 of 376) of the back-of-book indexes used cross-references. A cross-reference was "[a] direction from one heading or subheading to another," and they usually were 'see' or 'see also' references. They occurred more often in the social sciences than in the other two major fields. As Table 2 shows, the differences among the three major subject fields were not very large. It was much more likely for cross-references to occur in subject/general indexes than in name indexes.

A large fraction (87%) of subject/general indexes in the social sciences had at least one cross-reference each.

The best of the social sciences seems to have been class HQ, covering the family, marriage, and women: every one of eight HQ indexes examined had cross-references.

How many cross-references were used?

The average index had 32 cross-references. In science/technology the mean was 40 cross-references per index, far greater than in either of the other two major fields (see Table 2). I wondered if the high frequency of cross-references in science/technology was due to the size of books in this field: the average science/technology book had at least 80 more pages than the average book in any of the other three areas. So, one also can ask how many cross-references there were per 1,000 pages of text in the book. This calculation confirmed that cross-references occurred more frequently in science/technology than in arts/humanities or the social sciences. For every 1,000 pages of text in a science/technology book there were 99 cross-references in the subject/general index, whereas the average for all subject areas and types of indexes was 96 per 1,000 pages (see Table 2). The subfield of science/technology with the greatest ratio of cross-references to pages of text was RA, public aspects of medicine, where four books averaged 256 cross-references per 1,000 pages.

The per-page ratio showed cross-references in subject/general indexes to be least frequent in the arts/humanities, with 89 cross-references per 1,000 pages; yet name indexes in arts/humanities had the highest rate of cross-references in all the subject areas: 75 references per 1,000 pages.

The ratio of cross-references to text pages was especially low in the seven PS (American literature) and the 13 PR (English literature) books, which respectively averaged only 19 and 39 cross-references per 1,000 pages.
What types of cross-references were used?
In the 300 books with cross-references, the average index had 17 'see' references, 14 'see also' references, and one other cross-reference. References of the 'other' type included: 'also see', 'and see', 'and see under', 'cf', 'cf also', 'ie', 'see above', 'see also above', 'see also below', 'see also under', 'see below', and 'see under'.

'See's were especially common in science/technology and in class Z (see Table 2). In science/technology, the average index with references had 21 'see' and 17 'see also' references. In both arts/humanities and the social sciences, the number of 'see' references was about the same as the number of 'see also'.

If an index used only 'see' or only 'see only' references, which was it likely to be?
The 'see' reference was dominant not only in science/technology but in all four subject fields. There was a striking pattern in the exclusivity of these references, where an index uses only one kind of cross-reference, for example all 'see's. Of the 300 books with cross-references, 53 (18%) used 'see' exclusively. Only 19 indexes (6%) used 'see also' exclusively (see Table 2). 'See also' is a more complex type of reference than 'see'. Perhaps that is why the British Standards Institution standard BS 3700:1988 devotes one and a half columns to 'see also' but only one-third of a column to 'see' references. 12

If an index used both 'see' and 'see also' references, were they used about equally?
Again the answer points to the dominance of 'see'. In the 218 indexes that had both types of cross-references, the mean ratio was 3.8 to 1. In one-third (74 of 218) of these indexes, the 'see' references occurred at least twice as often as the 'see also' references. This dominance of 'see' was most visible in the six name indexes that had both types of references. The great need to send readers to the preferred form of a name caused the ratio of 'see' to 'see also' references to be very high: 13.5 to 1 (see Table 3).

Was there any situation in which 'see also' was the dominant reference?
In 56 indexes the ratio of 'see also' to 'see' references was large, at least 2 to 1. Seventeen of these indexes were in the arts/humanities: interesting, because there were only 36 arts/humanities indexes having both types of cross-references. It was not clear why arts/humanities was the one place 'see also' was sometimes used noticeably more than 'see'. Despite the dominance of 'see also' in these 17 indexes, the mean ratio of 'see' to 'see also' in the 36 arts/humanities indexes with both types of references still was slightly in favour of 'see', by 1.6 to 1.

The emphasis on either 'see' or 'see also'
The typical back-of-book index with cross-references had a clear preference either for 'see' or 'see also'. It was

<table>
<thead>
<tr>
<th>Subject/general indexes</th>
<th>Name indexes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AH</td>
<td>ST</td>
</tr>
<tr>
<td>No. of indexes</td>
<td>68</td>
<td>119</td>
</tr>
<tr>
<td>No. of indexes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>with cross-references</td>
<td>55</td>
<td>94</td>
</tr>
<tr>
<td>%</td>
<td>81</td>
<td>79</td>
</tr>
<tr>
<td>with exclusively 'see' cross-references</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>%</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>with exclusively 'see also' cross-references</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>%</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Mean No. of cross-references per index:* 'See'</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td>'See also'</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>40</td>
</tr>
<tr>
<td>Mean No. of cross-references per 1,000 text pages*</td>
<td>89</td>
<td>99</td>
</tr>
</tbody>
</table>

*Calculated only for the 300 indexes with cross-references.
unusual to find an index in which 'see's and 'see also's occurred at about the same frequency. Was this because most of the books and their users had a definite need especially for one type of reference, or because most of the indexes truly needed especially one type? In any case, 202 of 300 books with cross-references used one of the types of references either exclusively or at least twice as often as the other.

A sample of 2,247 cross-references

From the 300 back-of-book indexes that had cross-references I sampled the first on each page of each index. For each of these 2,247 cross-references, I examined the reference itself, the entry(ies) in the index to which it was referring, and if necessary, pages in the text represented by the reference. Then I answered two questions about each of these cross-references: (1) What type of cross-reference was it? The answer was usually 'see' or 'see also'. (2) What was its purpose? What kind of link was it making between entries in the index? The answer usually was one of five categories suggested by indexing standard BS 3700:1988. For example, the standard suggests having references 'from a class to its individual members'. But I categorized under class membership even those references that led users from an individual member to the class.

To these five categories of reference links, I added two others:

Blind/unknown: covered references for which either I could not find the index entry being referred to or I could not understand the purpose of the link.

Other: included all references that did not fit into any of the other categories above.

The results of this categorization are in Table 4.

Subjective as the categorizations were, I tried to reduce subjectivity by relying as much as possible on the author and indexer. For example, the author often gave clues as to whether two terms linked by the indexer were synonymous. The indexer might have created the reference: 'Head man: see also Slave drivers'. On one or more of the pages referred to by these entries, I might find the author saying: 'The head man, sometimes known as a slave driver...'. This was a clue for synonymy.
CROSS-REFERENCES IN BACK-OF-BOOK INDEXES

What were the most common links for the 2,247 cross-references?

The most common links occurred in the following categories: synonyms, class membership, activities, and 'other'.

Most common was the link between terms that were synonymous in some way. As Table 4 indicates, the synonyms category accounted for almost half the cross-references. A typical reference was 'Animals, study of: see Zoology'. Various relationships were represented in the 1,024 synonym cross-references.

The analysis of links provided a partial explanation for the dominance of the 'see' reference—about 80% of the synonym links. Synonyms accounted for more cross-references than any of the other links. Therefore, 'see' was dominant despite that fact that 'see also' occurred at least as often as 'see' in each of the other six categories of links.

After synonyms, the next most common type of link occurred in the 434 references that linked a class with a member of the class. I felt that links between broader and narrower terms in the same class also fell into this category. Examples were 'Christianity: see also Religion' and 'Safety: see also Workplace safety'.

Third were 368 activities links. There were varieties of these, such as, 'Roads: see also Transportation', which was a link between an agent ('Roads') and its activity or product ('Transportation').

The fourth most commonly occurring of the seven categories of links was the miscellaneous 'other' category. Many types of cross-references seemed not to fit neatly into the links suggested by standard BS 3700:1988.

Were there differences in cross-reference links among the four subject fields?

There were few differences from subject to subject, and almost no differences between the arts/humanities and the social sciences. The percentage breakdown of the seven categories of links in Table 4 was nearly identical for these two fields. Science/technology and the Z category were somewhat different in that their use of the class membership link was greater than for the other two fields. In science/technology class membership occurred so frequently because of the many links between a substance or an organism and the class to which it belonged.

Did the standards mention all the possible links for which cross-references were used?

No, that is not the nature of indexing standards. Nevertheless, the variety of links found in the 300 indexes with cross-references was almost bewildering. This reflected the artistic nature of indexing and the individual differences that exist from book to book and from indexer to indexer.

To account for the variety of links, I analyzed them more finely than the list in Table 4. Many finer subcategories of links were suggested by Borko and Bernier, Cleveland and Cleveland, as well as by standard BS 3700:1988.

Table 4. Analysis of 2,247 cross-reference links

<table>
<thead>
<tr>
<th>Subject categories</th>
<th>Types of cross-references</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>'See'</td>
<td>'See also'</td>
</tr>
<tr>
<td>Synonyms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>43%</td>
<td>49%</td>
</tr>
<tr>
<td>Class membership</td>
<td>62%</td>
<td>188%</td>
</tr>
<tr>
<td>%</td>
<td>15%</td>
<td>24%</td>
</tr>
<tr>
<td>Activities</td>
<td>75%</td>
<td>84%</td>
</tr>
<tr>
<td>%</td>
<td>18%</td>
<td>10%</td>
</tr>
<tr>
<td>Whole/part</td>
<td>31%</td>
<td>35%</td>
</tr>
<tr>
<td>%</td>
<td>8%</td>
<td>4%</td>
</tr>
<tr>
<td>Field/objects/applications</td>
<td>16%</td>
<td>24%</td>
</tr>
<tr>
<td>%</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td>Blind/unknown</td>
<td>6%</td>
<td>12%</td>
</tr>
<tr>
<td>%</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>Other</td>
<td>44%</td>
<td>62%</td>
</tr>
<tr>
<td>%</td>
<td>11%</td>
<td>8%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>11%</td>
<td>796%</td>
</tr>
</tbody>
</table>


182 The Indexer Vol. 17 No. 3 April 1991
CROSS-REFERENCES IN BACK-OF-BOOK INDEXES

Looking more closely at the links for synonym

One-third (315) of the 1,024 synonym references were simply links between words or phrases with similar meanings. The link usually used ‘see’ to send the user to the preferred synonym.

The other two-thirds of the synonym references were links between terms that had similar meanings and differed in physical format. These included name variations, abbreviations, word-order differences, translations, and chemical synonyms.

There were 339 synonym links between two forms of a name. The names were usually names of people, but also could be proper names for places or organizations. The link was straightforward, such as ‘Petersburg: see St Petersburg’. The importance of names in literature makes it not surprising that 59% of the arts/humanities synonym links were of this kind, and only 5% of the science/technology synonym links were name links.

There were 142 synonym links between a term and its abbreviation or acronym. An example was ‘H-D method: see Hypothetico-deductive method’. Not one of these abbreviation links occurred in the sample taken from the 64 arts/humanities indexes. Abbreviations were much more common in science/technology and social sciences books than in arts/humanities.

There were 132 synonym links between phrases that differed in word order; most were links between direct and inverted forms of a phrase. These formats were almost equally likely to be preferred. For the direct term as the preferred term, there were 34 links, such as ‘Clouds, stratus: see Stratus clouds’. For the indirect term as the preferred term, there were 46 links, such as ‘Woman as a class: see Class, woman as a’. Almost all the links between terms differing in word order were in science/technology indexes.

There also were synonym links between a term and its translation (‘Makhorka: see also Tobacco’); chemical synonyms (‘A-23187: see Calcimycin’); antonyms (‘Generalities: see also Specifics’); a colloquialism or piece of jargon and its formal expression (‘Indies: see Individual record companies’); and spelling variations (‘Cabalah: see Kabbalah’).

Looking more closely at the links for class membership

In the class membership links, I finely examined only their direction. First, a cross-reference could send a reader from a specific, narrower term to a broader term that named the class in which the narrower term resided. One-third of the 434 class membership references did this, as in ‘Abelian group: see Group’. ‘See’ and ‘see also’ references were equally common here. Second, a cross-reference could send a reader from a broader term to a narrower or more specific term that named a member of the class. Two-thirds of the 434 class membership references did this, usually via a ‘see also’ reference, as in: ‘Survival: see also Absolute survival’.

Looking more closely at the links for activities

These references linked activities, their agents, and the products (or results or receivers) of the activities. ‘See also’ was much preferred to ‘see’ in these situations.

About half (175) of the 368 activities references linked the agent of an activity with its product. I made some very subjective decisions here; others might have categorized some of these links differently. An example is ‘Ku Klux Klan: see also Terrorism’, where I felt the author saw terrorism as the product of Ku Klux Klan activity. Also included in this category were links between work and its creator, as in ‘DeLillo, Don: see Names of individual works’.

There were 75 activities references in which the link was between an activity and its product. Examples include the links of a product and its use, such as ‘Books: see also Reading’; as well as between a cause and its effect, as in ‘Characterizing students: see Communications, judgmental’. The latter is another indication of the importance of the author in my categorizations. This is a cause-effect link because the author indicated that characterizing students can lead to judgmental communications.

There were also activities references that linked an activity or participation in it and its agent or participant. Examples were ‘Election of 1860: see also Lincoln, Abraham’ and ‘Epicureanism: see also Lucretius’.

There were even activities references that linked an activity or agent with the place or time of the activity, as in ‘Greeks: see also Greece’ and ‘Establishment: see also Sixties, the’.

Looking more closely at some of the other links

Less than one-tenth (168) of all the 2,247 sampled cross-references made links that did not fit well into any of the above categories or subcategories. These included links between: a thing and its characteristics (‘Validity: see Tests’); two members of a class; a person and his/her relative; an eponymous thing and the person for which it was named; and a group of people and their creed or philosophy.

Summary and comments: the positive, the negative and the surprising

The most positive finding was that 80 per cent of the back-of-book indexes employed cross-references. This was especially encouraging when we consider that the indexes probably were created by a variety of indexers, including authors as well as professional indexers. Cross-references were common in all three of the major subject areas studied: arts/humanities, science/technology, and
CROSS-REFERENCES IN BACK-OF-BOOK INDEXES

the social sciences. The rate of cross-referencing was not vastly different in these three fields. For the indexes that had cross-references, the average contained 96 cross-references per 1,000 pages of text.

Also encouraging was that in sampling over 2,000 cross-references, I saw few errors. In only seven cases did the referenced entry not exist. Almost always, on turning to the page indicated by an entry, I was able to find the item referred to in the index.

It was good to see that indexers and the writers of standards and indexing textbooks agree on the major reasons for cross-reference links. Linking a pair of synonyms, or a class member with its class, or an activity with its agent or product are accepted as common reasons to provide cross-references. It was also good to see that indexers can find many other reasons to create a cross-reference link. From linking entries for two people because they are relatives to linking an entry for a group with the beliefs of that group, indexers can be very creative in helping readers.

On the negative side was the paucity of cross-references in name indexes. I examined only 32 name indexes, and many of them seemed appropriately simple and straightforward. Would cross-references really be needed in most name indexes? Of course it depends on the index, but the findings did show that the linking of various forms of names was a frequent purpose of 'see' references in the subject/general indexes. If cross-references are useful for names in a general index, then they also ought to be important in a name index.

Also disappointing was confusion between 'see' and 'see also' in the few indexes. Several times the indexer reversed the role of these references. An example of 'see' used as a 'see also' is: 'Cats: 34, 45: see Feline diabetes'.

The surprising result was the dominance of one of the two common cross-references, either 'see' or 'see also'. The dominance was not immediately apparent, for the average book rarely had about the same number of 'see's and 'see also's. It was common for one type of cross-reference to outnumber the other by at least two to one. The only explanation I can offer is that the content of a given book, or the nature of its readers, or the tendencies of an indexer made one type of cross-reference much more appropriate than the other.

In these 300 indexes with cross-references, the dominant reference was usually 'see'. Perhaps 'see' is perceived as easier to use (by the reader and by the indexer) than 'see also'. Directing readers to preferred terms may be perceived as having priority over suggesting other index entries to them.

This investigation did not examine alternatives for cross-references, such as multiple entries. Placing the same page locators with two synonymous entries can be an alternative to linking the entries with a cross-reference. There was creativity and variety in the 376 back-of-book indexes examined here, and one would expect that many of them employed multiple entries, too.

References

Virgil Diodato is Assistant Professor at the University of Wisconsin-Milwaukee School of Library and Information Science.