The Influence of Existing Consistency Measures on the Relationship between Indexing Consistency and Exhaustivity

Abstract

Consistency studies have discussed the relationship between indexing consistency and exhaustivity, and it is commonly accepted that higher exhaustivity results in lower indexing consistency. However, this issue has been oversimplified, and previous studies contain significant misinterpretations. The aim of this study is to investigate the relationship between consistency and exhaustivity based on a large sample and to analyse the misinterpretations in earlier studies. A sample of 3,307 monographs, i.e. 6,614 records was drawn from two Chinese bibliographic catalogues. Indexing consistency was measured using two formulae which were popular in previous indexing consistency studies. A relatively high level of consistency was found (64.21% according to the first formula, 70.71% according to the second). Regarding the relationship between consistency and exhaustivity, it was found that when two indexers had identical exhaustivity, indexing consistency was substantially high. On the contrary, when they had different levels of exhaustivity, consistency was significantly low. It was inevitable with the use of the two formulae. Moreover, a detailed discussion was conducted to analyse the misinterpretations in previous studies.

Introduction

Ideally, if two indexers use the same thesaurus or classification system to index the same book, they are supposed to assign the same index terms or class numbers. In practice, indexers are not always consistent with each other, because subject indexing is essentially a subjective process. Since indexing consistency is an acceptable indicator of indexing quality, the increasing interest in enhancing information retrieval effectiveness has heightened the need for improving indexing consistency between indexers. Many factors can affect indexing consistency, among which indexing exhaustivity is the most important one and has been most extensively explored in former consistency studies. Considering the relationship between indexing exhaustivity and consistency, it is widely believed that higher exhaustivity leads to lower indexing consistency. However, the relationship between indexing consistency and exhaustivity is not as straightforward as it might seem. Actually, this issue has been oversimplified, and previous studies contain significant misinterpretations. The aim of this study is to investigate the relationship between consistency and exhaustivity based on a large sample and to analyse the misinterpretations in earlier studies.

Methodology

A sample of 3,307 monographs (i.e. 6,614 records) was selected from two Chinese bibliographical databases. One is the catalogue of the National Library of China (NLC), the other one is the catalogue of the China Academic Library & Information System (CALIS). NLC is the biggest public library system in China, while CALIS is the biggest academic library system. Both of the library systems utilize the Chinese Thesaurus (CT) as an indexing tool, which makes it convenient to comparing indexing results.

In earlier studies, the following two formulae were extensively used to calculate indexing consistency:

\[ CP1 = \frac{c}{a+b-c} \] (Hooper, 1965, reported by Leonard, 1977)
\[ CP2 = \frac{2c}{a+b} \] (Rolling, 1981)
where \( a \) denotes the number of terms assigned by one indexer, \( b \) denotes the number of terms assigned by a second indexer, \( c \) denotes the number of terms commonly assigned by the two indexers. In the interest of comparing with previous studies, both formulae are used for calculating indexing consistency in this study.

**Results**

The consistency of index terms between NLC and CALIS is calculated with the use of the two formulae described above. According to Hooper’s formula, the overall consistency of index terms is 64.21%, while according to Rolling’s formula it is 70.71%. A total of 18,182 index terms are assigned to the 6,614 records, i.e. 3,307 books, averaging 2.75 terms per record (Sd.1.20). The indexers of NLC totally assign 8,999 terms to these 3,307 books, averaging 2.72 terms per book (Sd. 1.13, std error mean 0.20.); while the indexers of CALIS totally assign 9,183 terms, averaging 2.78 terms per book (Sd. 1.27, std error mean 0.22). A t-test shows no significant difference between the two catalogues with regard to the number of index terms. That is to say, the two catalogues have identical level of indexing exhaustivity.

**Figure 1.** Consistency and total number of terms assigned by two indexers per book

The data shows whether two indexers assign equal number of index terms to the same book profoundly influences indexing consistency. With reference to Figure 1 we can see that when the total number of index terms assigned by two indexers to the same book is an even number, the indexing consistency is considerably high. In contrast, when the total number is an odd number, the indexing consistency is obviously low. Therefore, we should discuss the relationship between consistency and exhaustivity under two different conditions. Say condition 1 stands for the situation when the total number of terms assigned by two indexers to the same book is an odd number; condition 2 stands for the situation when the total number of terms assigned by two indexers to the same book is an even number.

Under condition 1, there is a positive correlation between consistency and exhaustivity. It indicates that the more terms two indexers totally assign to a book, the more likely they assign the same terms. Under condition 2, the relationship between consistency and the total number of terms assigned is a little more complicated. The consistency reaches a peak of 93.05% when two terms are totally assigned to a book by two indexers, then it
declines sharply when four terms are assigned. But, after that it rises gradually, although there is a slight drop at the level of ten terms (see Figure 1). That is to say, in both cases consistency rises as the total number of terms assigned by two indexers per book increases, except for the extreme value 93.05%, which can be regarded as an outlier. In conclusion, consistency tends to be higher, when more terms are assigned to a book by two indexers.

**Analysis**

The experimental data has shown that when two indexers have identical level of exhaustivity, the indexing consistency is considerably high. Furthermore, the smaller differences in the number of terms assigned by two indexers lead to a greater probability of achieving higher consistency (see Figure 2). The reason is that when two indexers assign unequal number of terms to the same book (the total number is an odd number), there is no probability of achieving 100% consistency. Contrarily, when two indexers assign equal number of terms (the total number is an even number), there is a great probability of reaching 100% consistency.

![Figure 2. The smaller differences in the number of terms assigned by two indexers, the higher consistency can be achieved](image)

Actually, it is inevitable with the use of the two formulae mentioned above, because the formulae used increase the probability of a high score if two indexers assign an equal number of terms to a document rather than an odd number. It can be clearly explained by the following examples. In these examples, the consistency scores are calculated based on Hooper’s formula.

Example 1: If indexer A assigns 10 terms to a document, while indexer B assigns 10 terms to this document as well; then the number of commonly assigned terms \(c\) can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. According to Hooper’s formula, the consistency scores can be 0, 1/19, 2/18, 3/17, 4/16, 5/15, 6/14, 7/13, 8/12, 9/11, 10/10. Assume that it is a discrete uniform distribution, and then the probability of each possible outcome is 1/11. The expected value of the consistency can be calculated by using the following function:

\[
E(\text{consistency}) = \sum_{c=0}^{10} \frac{1}{11} \times \frac{c}{20-c}
\]

In this case, the expected value is 0.40.
Example 2: If indexer A assigns 5 terms to a document, while indexer B assigns 15 terms; then the number of commonly assigned terms $c$ can be 0, 1, 2, 3, 4, 5. The consistency scores can be 0, 1/19, 2/18, 3/17, 4/16, 5/15. In this case, the expected value of the consistency is: 0.15.

Example 3: If indexer A assigns 1 term to a document, while indexer B assigns 19 terms; then the number of commonly assigned terms $c$ can only be 0 or 1. The consistency scores can be 0 or 1/19. In this case, the expected value of the consistency is: 0.03.

It is very clear that if the total number of terms assigned by indexer A and B is constant ($a+b$ is constant), the possibility that the two indexers are consistent with each other is the highest, when they assign equal number of terms. The smaller differences are between the numbers of terms they assign, the greater probability they get to achieve high consistency.

**Figure 3.** Consistency decreases with the increase of the total number of terms assigned to a book by two indexers, when the number of terms commonly assigned is constant.

Another interesting phenomenon is that when the number of terms commonly assigned to a book by two indexers is constant ($c$ is constant), the more terms the two indexers totally assign to this book, and the less consistent they are, which can be shown by the curves in Figure 3. As the total number of terms increases, the distances between these curves become smaller. That is to say, the more terms the two indexers totally assign to a book, the smaller are the differences in the consistency scores, no matter how many terms they commonly assign.

Figure 3 seems to contradict the conclusion that consistency tends to be higher, when more terms are assigned to a book by two indexers. It is essential to bear in mind that the curves in Figure 3 show the relationship between consistency and total number of terms assigned by two indexers per book, when the number of terms commonly assigned by two indexers is constant. However, in practice, the number of terms commonly assigned by two indexers is not constant. In fact, it varies with the total number of terms assigned by two indexers. The experimental data show that the more terms two indexers totally assign to a book, the more terms they commonly assign. Theoretically, it is also true, because when the total number of terms assigned by two indexers increases, the probability that they use the same terms is larger. In the light of Hooper’s formula, $c = \{0, 1,$
2, \ldots , (a+b)/2}, where \( a+b \) is even; or \( c = \{0, 1, 2, \ldots, (a+b-1)/2\} \), where \((a+b)\) is odd. If we calculate the expected value of \( c \), we can find that the larger \( a+b \) is, the larger the expected value of \( c \) is.

The experimental data also show that the more common terms two indexers assign to a book, the more consistent they are (see Figure 4). However, it is not a linear relationship between the number of terms commonly assigned and indexing consistency. As the number of terms commonly assigned increases, indexing consistency infinitely approaches 100%, but will never reach it.

**Figure 4.** Relationship between the number of terms commonly assigned to a book by two indexers and indexing consistency

Since the more terms two indexers assign to a book, the more terms are in common, and meanwhile the more terms are in common, the higher consistency can be achieved, we can say that it is inevitable that two indexers are more consistent with each other when they assign more terms to a book.

**Discussion**

There is a common accepted statement about the relationship between indexing consistency and exhaustivity in this field: consistency drops as indexing exhaustivity increases. This belief appears to be supported by Fried & Prevel (1966, reported by Leonard 1977), Reich & Biever (1991), Sievert (1991) and Shoham (2001). However, this is not always the case, because the findings of the studies conducted by Harris (1966, reported by Leonard 1977), Leonard (1975, reported by Leonard 1977) contradict it.

In Reich and Biever’s (1991) study, they found that for a sample of articles indexed with an average of 8–9 thesaurus terms, the consistency was 24%; it was 45% for a sample having an average of 5–6 thesaurus terms. However, if we examine Reich and Biever’s results closely, we can find that it is highly debatable to draw such a conclusion based on their data, because about 16.67% of the descriptors were identical to terms in title in relation to sample #1, while about 46.55% in relation to sample #2. Theoretically, when indexers extract index terms directly from titles, they can be more consistent. Thus, the consistency was to some extent determined by how many descriptors were identical to terms in title. In this case, we cannot be sure whether the difference in consistency between sample #1 and sample #2 was resulted by different levels of exhaustivity or by different numbers of descriptors which were identical to terms in title.
Sievert (1991) compared the consistency of main headings (50.39%) with that of sub-headings (47.89%), as well as comparing the consistency of descriptors (47.27%) with that of identifiers (32.83%). And he believed that the differences were resulted by the different levels of indexing depth. However, the properties of these different categories of index terms haven’t been taken into account in his research. Generally, main headings are used to represent major topics of a document, while subheadings are used to represent minor topics or specific aspects of a topic. Some evidence in the literature showed that indexers were more likely to agree on central aspects of a work than on peripheral themes (Lancaster, 1998; Funk et al., 1983; Iivonen, 1990). Besides, according to Sievert, in his sample, the descriptors were assigned from a small controlled vocabulary, while the identifiers were assigned from natural language. As it is known, controlled indexing has a beneficial effect on indexing consistency (Markey, 1984). Therefore, it is not convincing to say that higher exhaustivity results in lower consistency based on the findings of Sievert’s study.

Shoham and Kedar (2001) have made a considerable improvement over previous studies. They directly calculated the correlation between indexing consistency and number of terms, and found that the Pearson correlation coefficient was between -0.28 and -0.36 in different cases. Their findings indicated that there was a negative correlation between consistency and exhaustivity, although the strength of the association between consistency and number of terms was not substantially strong.

As mentioned above, some researchers took a contrary position. Harris (1966) maintained that ‘consistency (% overlap) does not seem to change much with depth… the graph (% consistency vs. depth) shows an almost horizontal line.’ Leonard (1977) also claimed that depth of indexing had no apparent effect on indexing consistency. Actually, whether two indexers assign equal number of index terms to the same book, i.e., whether they have identical indexing depth has profound effect on indexing consistency. Some evidence can be found in the literature.

Reich and Biever (1991) realized that different indexing depths could result in low consistency by saying: “the difference in the number of terms assigned to each document obviously contributes to the observed low index match rate of 27 percent.” If we examine Tonta’s study (1991), we can find the same reason for the low consistency (16% for exact matches and 36% for both exact and partial matches). According to his report, LC cataloguers assigned 282 subject headings for 82 items while BL cataloguers assigned 127. The difference in indexing depth between the two catalogues was significant. Stubbs’ experiment (1999) also provides some evidence. In the experiment, seven students indexed two printed manuals on library and information science twice. At the first indexing, indexer a, b, c and d were very selective (2–3 terms), indexer g was very exhaustive (30 terms), and indexer e and f applied 8 and 7 terms, respectively. But, at the second indexing, they had almost the identical exhaustivity (a 21, b 18, c 21, d 17, e 21, f 22, and g 25). Then, the mean consistency increased from 29% to 60%. There might be some other reasons for the increase. Nonetheless, we cannot deny that identical exhaustivity also contributed to the increase in consistency.

**Conclusion**

To sum up, the relationship between indexing consistency and exhaustivity is complicated. Whether two indexers assign equal number of index terms to a document affects profoundly indexing consistency. Although identical exhaustivity is not a sufficient condition
for high consistency, it is a necessary one. The analysis based on the experimental data of this study has shown that the widely accepted belief is actually wrong. There is no doubt that exhaustive indexing is beneficial rather than detrimental to indexing consistency. In the new networked environment, new optimal level of indexing exhaustivity needs to be determined. Moreover, further studies are needed to understand the practical meanings of various consistency measures. Maybe a new consistency measure is needed, so that we can discuss the relationship between indexing consistency and exhaustivity without the influence of the measures used.

References