Humboldt-Universität zu Berlin

Institut für Bibliotheks- und Informationswissenschaft

Dissertation

Indexing Consistency between Online Catalogues

Zur Erlangern der Doktorwürde

Philosophische Fakultät I.

Xu Chen

Dekan der Philosophische Fakultät I.: Prof. Dr. Michael Borgolte

Gutachter: 1. Prof. Dr. Walther Umstätter

2. Prof. Dr. Robert Funk

eingereicht: 15.10.2007

Datum der Promotion: 14.05.2008

Zusammenfassung

In der globalen Online-Umgebung stellen viele bibliographische Dienstleistungen integrierten Zugang zu unterschiedlichen internetbasierten OPACs zur Verfügung. In solch einer Umgebung erwarten Benutzer mehr Übereinstimmungen innerhalb und zwischen den Systemen zu sehen. Ein konsistenter Indexierungsdienst kann die Vorhersagbarkeit verbessern und bessere Suchergebnisse erzielen. In der Praxis stimmen die Indexiere mit einander nicht immer überein, da Inhaltliche Erschließung im Wesentlichen ein subjektiver Prozess ist.

Zweck dieser Studie ist, die Indexierungskonsistenz zwischen Systemen zu untersuchen und herauszufinden, ob diese auch in der neuen, vernetzten Umgebung enttäuschende Ergebnisse liefert. Währenddessen werden einige Faktoren, die weitgehend in vorherigen Studien besprochen werden, in dieser Studie untersucht, um es sehen zu können, ob sich die Verhältnisse zwischen diesen Faktoren und der Indexierungskonsistenz geändert haben. Wichtigstes Ziel dieser Studie ist, die Gründe für die Inkonsistenzen herauszufinden, damit sinnvolle Vorschläge gemacht werden können, um die Indexierungskonsistenz zu verbessern.

Eine Auswahl von 3307 Monographien wurde aus zwei chinesischen bibliographischen Katalogen gewählt. Die Indexierungskonsistenz wurde mit zwei Formeln gemessen, die in vorherigen Indexierungskonsistenz Studien häufig verwendet wurden. Ein verhältnismäßig hohes Niveau an Übereinstimmungen wurde gefunden. Nach Hooper's Formel war die durchschnittliche Indexierungskonsistenz für Indexterme 64,2% und für Klassennummern 61,6%. Nach Rolling's Formel war sie für Indexterme 70,7% und für Klassennummern 63,4%.

Mehrere Faktoren, die die Indexierungskonsistenz beeinflussen, wurden untersucht: (1) Indexierungsbereite; (2) Indexierungsspezifizität; (3) Länge der Monographien; (4) Kategorie der Indexierungssprache; (5) Sachgebiet der Monographien; (6) Entwicklung von Disziplinen; (7) Struktur des Thesaurus oder der Klassifikation; (8) Erscheinungsjahr. Die Ergebnisse von dieser Studie zeigten, dass in vorherigen Studien die Beziehungen zwischen die Faktoren und die Indexierungskonsistenz übervereinfacht oder missinterpretiert wurden.

Gründe für die Inkonsistenzen wurden ebenfalls analysiert. Die Analyse ergab: (1) den Indexieren mangelt es an Fachwissen, Vertrautheit mit den Indexierungssprachen und den Indexierungsregeln, so dass viele Inkonsistenzen verursacht wurden; (2) der Mangel an vereinheitlichten oder präzisen Regeln brachte ebenfalls Inkonsistenzen hervor; (3) verzögerte Überarbeitungen der Indexierungssprachen, Mangel an terminologischer Kontrolle, zu wenige Erläuterungen und "siehe auch" Referenzen, sowie die hohe semantische Freiheit bei der Auswahl von Deskriptoren oder Klassen, verursachten Inkonsistenzen.

Schlagwörter:

Indexierungskonsistenz, Thesauri, Klassifikationen, Indexierungssprache, Indexierungsbereite, Indexierungsspezifizität, Indexierungsregeln, Inhaltliche Erschließung

Abstract

In the global online environment, many bibliographic services provide integrated access to different web-based OPACs. In such an environment, users expect to see more consistency within and between systems. A consistent indexing service can improve predictability, helping users to acquaint themselves with the indexing practices and to achieve better retrieval results. In practice, indexers are not always consistent with each other, because subject indexing is essentially a subjective process.

The purpose of this study is to investigate the indexing consistency between systems and to find out whether it is still frustrated in the new networked environment. Meanwhile, some factors which are widely discussed in former studies will be examined in this study, so that we can see whether the relationships between these factors and indexing consistency have changed. The most important aim of this study is to find out the reasons for inconsistencies, so that some reasonable suggestions can be made to improve indexing consistency.

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. According to Hooper's formula, the average consistency for index terms was 64.2% and for class numbers 61.6%. According to Rolling's formula, for index terms it was 70.7%, for class numbers 63.4%.

Several factors affecting indexing consistency were examined: (1) exhaustivity of indexing; (2) specificity; (3) length of monographs indexed; (4) category of indexing languages; (5) subject area of monographs indexed; (6) development of disciplines; (7) structure of vocabulary; (8) year of publication. The results of this study showed that some issues have been oversimplified and some significant misinterpretations were contained in previous studies.

The reasons for inconsistencies were also analyzed. The analysis revealed that: (1) indexers' lack of subject knowledge, their unfamiliarity with indexing languages and indexing rules led to a lot of inconsistencies; (2) the lack of unified or detailed indexing policies brought about inconsistencies as well; (3) delayed revision of indexing languages, lack of vocabulary control, shortage of scope notes and “see also” reference notes, and high semantic freedom by term or class choosing also caused inconsistencies.

Keywords:

Subject indexing, indexing consistency, thesauri, classifications, exhaustivity, specificity, indexing languages, indexing policies

Dedication

To my unborn daughter, who brought me the hope of life,
and to my husband, who taught me the meaning of love.

Acknowledgements

Many people have contributed either directly or indirectly to this study. It is my pleasure now to acknowledge their support and contribution.

First and foremost, I want to express my sincerest gratitude to my advisor at the Institute of Library and Information Science at Humboldt University in Berlin, Professor Dr. Walther Umstätter, who provided me a lot of insightful comments and suggestions. He gave me encouragement and advice when I encountered difficulties

My special thanks go to Professor Zhanghua Ma at the department of Library and Information Science at Peking University, who encouraged me to come to Germany and continue my study in this field.

Acknowledgements are also due to Michael Heinz at the Institute of Library and Information Science at Humboldt University in Berlin for his valuable advice and help with statistics.

I own many thanks to my friend, Amy Zhang, who proofread the text.

Lastly, no words can express my deep gratitude to my husband, Yuning Yang, who supported me not only financially but also psychologically in the past several years. Every time when I was frustrated or lost my faith, he inspired me with hope.

Table of Contents

Abbreviations

ANOVA: ANalysis Of VAriance

CALIS: China Academic Library & Information System

CLC: Chinese Library Classification

CT: Chinese Thesaurus

DDC: Dewey Decimal Classification

Df: degree of freedom

LC: Library of Congress

MARC: Machine Readable Cataloging

NLC: National Library of China

OCLC: Online Computer Library Center

ODLIS: Online Dictionary for Library and Information Science

OLCC: Online Library Cataloging Center

OPAC: Online Public Access Catalog

Sd.: Standard Deviation

List of figures

List of tables

1 Introduction

Many recent studies have focused on users' information retrieval behavior and shown that users of online catalogues prefer subject searching. However, in many cases, the users are frustrated by retrieving zero hits or more than one thousand hits. The failure of subject searching to some extent is resulted by low indexing quality. Traditionally, indexing consistency is considered as an acceptable indicator of indexing quality. Indexing consistency refers to "the extent to which agreement exists on the terms to be used to index some document" (Lancaster, 1998). Ideally, if two indexers use the same thesaurus or classification system to index the same document, 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.

Nowadays, in the global online environment, many bibliographic services provide integrated access to different web-based OPACs. In such an environment, users expect to see more consistencies within and between systems (Fattahi, 1998). Theoretically, a consistent indexing service can improve predictability, helping users to acquaint themselves with the indexing practices and to achieve better retrieval results. In practice, the result of a study also indicates that indexing consistency has a positive influence on retrieval effectiveness (Leonard, 1975). That is, higher consistency is of great benefit for retrieval effectiveness. The increasing interest in enhancing information retrieval effectiveness has heightened the need for improving indexing consistency between systems.

The issue of indexing consistency has been extensively studied in recent years. The results of many studies have shown low indexing consistency, which has made some researchers draw the conclusion that indexing consistency levels could rarely cross the 50% threshold on a standard scale of 0% to 100% (Huton, 1998; Bate, 1986). Some researchers believe that "the indexers differ considerably in their judgment as to which terms reflect the contents of the document most adequately" (Zunde & Dexter, 1969). Some other researchers even challenge the necessity of spending much time to determine whether the subject analyses are consistent, and challenge the theory of collocation being the ultimate goal of subject cataloguing. They suggest that cataloguers can be more prone to accept variations in subject choices in member copy (Gregor & Mandel, 1991). In some library schools there is even a debate of whether it is necessary to assign subject headings any more (Mann, 1997).

However, so far, most previous studies are experimental or have been confined to indexing consistency within systems. The indexers in these experiments are usually novices or even users and they perform the indexing task without any controlled vocabularies. In most cases, the sample sizes are limited to 100 items because of the shortage of resources. Therefore, the results of these studies are not representative for the real life cataloguing, especially when the situation is changing in the networked environment, where new investigations are needed. Although the development of web-based OPACs and indexing languages has led to the hope of achieving high indexing consistency, the inconsistency problem is still apparent among indexers between systems.

The purpose of this study is to investigate the indexing consistency between systems and to find out whether it is still frustrating in the new networked environment. Meanwhile, some factors which are widely discussed in former studies will be examined in this study, so that we can see whether the relationships between these factors and indexing consistency have changed. Actually, in previous studies, some issues have been oversimplified and some significant misinterpretations are contained. In this study, these factors will be investigated based on a large sample, which is drawn directly from the real cataloguing environment of two large bibliographic databases. On this basis, the misinterpretations in earlier studies will be discussed. But, the most important aim of this study is to find out the reasons for inconsistencies, so that some reasonable suggestions can be made to improve indexing consistency.

The remainder of this thesis is divided into four chapters: First, how this investigation is conducted and how the data are analyzed are introduced in Chapter 2. In Chapter 3, the relationships between indexing consistency and some factors are closely examined on the basis of quantitative data. The reasons for inconsistencies are explored by thoroughly analyzing the contents of the sampled bibliographic records. After the in-depth quantitative and qualitative analyses, a detailed discussion is provided in

Chapter 4 to explain the results of this study and to analyze the misinterpretations in previous studies. In the final chapter (Chapter 5) the whole study is summarized and some conclusions are drawn. Then, some recommendations for improving indexing consistency and for further research are provided.

2 Material and methods

2.1 Introduction

This study represents an attempt to compare indexing consistency between two online catalogs. Since it happens frequently that a book is indexed in different catalogs, it is possible to collect pairs of bibliographical records and compare them. Theoretically, different indexers should assign the same class numbers and index terms to the same titles, if they use the same indexing languages. In practice, they cannot be totally consistent with each other.

The indexing languages involved in present study are introduced in Section 2.2. Purposeful sampling is utilized to collect the data, which is to say, the sample is drawn according to a set of criteria or a list of attributes. Detailed sampling procedures can be found in Section 2.3. Various consistency measures have been used in previous studies. They are introduced and compared in Section 2.4. After the data are collected, they should be compared and analyzed. Section 2.5 deals with some rules applied to the data analysis. Finally, Section 2.6 and Section 2.7 contain a brief introduction of the statistical analysis and the conceptual analysis involved in this study.

2.2 The two involved indexing languages

2.2.1 Chinese Library Classification

The Chinese Library Classification (CLC for short), which was first published in 1975, is a comprehensive classification scheme that is widely used in most Chinese libraries, information institutes and centres. It’s playing the role of a national standard in the complication and usage of Chinese classifications (Zeng, 1992). The second edition of the classification was published in 1980, and the third in 1990. In 1999, its fourth edition (the newest version) was published after an effective revision to three key classes--Finance (F), Radio electronics Tele-technology (TN), and Automation technology Computer Science (TP). The important thing is that the fourth edition is the first to be converted into a machine-readable form by a firm and is available on CD-ROM (CLC Editorial Board, 1999).

The CLC is a universal scheme that treats knowledge as a whole. The whole knowledge system is divided into 22 main classes that are denoted by a capital letter as follows:

A Marxism, Leninism, Maoism & Deng Xiaoping's Theory
B Philosophy and Religion
C Social Science
D Politics and law
E Military Science
F Economics
G Culture, Science and Education
H Linguistics
I Literature
J Art
K History and Geography
N Natural Science
O Mathematics, Physics and Chemistry
P Astronomy and Geoscience
Q Bioscience
R Medicine and Hygiene
S Agricultural Science
T Industrial Technology
U Transportation
V Aviation
X Environmental Science
Z Comprehensive books

These 22 main classes are organized under the following five categories:

- Marxism, Leninism, Maoism & Deng Xiaoping's Theory
- Philosophy and Religion
- Social sciences
- Nature sciences
- Comprehensive books

The CLC has been strongly impacted by the DDC. The scientific notations, structures and advanced compiling techniques have been mostly referenced from the DDC. But some modifications have been also made, e.g. the mixed notation and some devices, which increase the capacity of the CLC. In recent years the DDC has always tried to keep abreast with world changes. Comparatively, the CLC develops very slowly, especially at the following aspects: timely updates to classification numbers, the online version of the classification and assisting in the task of organizing the Internet resource (Hong and Zhang, 1996).

2.2.2 Chinese Thesaurus

The Chinese Thesaurus (CT for short) was first published with 91,158 preferred terms and 17,410 non-descriptors in 1980. It is the largest and most comprehensive thesaurus in China. It contains a substantial number of terms and deals with a wide range of disciplines and technologies (Zhang, 2004).

The first edition consisted of three massive parts and 10 volumes: social sciences and humanities (Part I, Volumes 1 – 2); natural sciences and technologies (Part II, Volumes 3 – 9); and appendixes (Part III, Volume 10). Each of the first two parts was composed of an alphabetical listing of terms, a hierarchical index, a subject category index, and an English-Chinese bilingual index. Appendixes included countries and regions of the world, geographical areas, and lists of organizations, agencies and important persons.

In 1991, the section of natural sciences and technologies was revised and enriched. 8,221 new descriptors were added and 5,424 terms were deleted. In 1996, a permuted index for natural sciences and technologies was added to the revised edition.

The Chinese Thesaurus has played a key role in the development of Chinese information indexing and retrieval, but it still has some problems. For instance, it lacks specificity, and it is inconsistent in quality and depth of descriptors among different disciplines. Moreover, it includes too many out-of-date terms and many terms that are approximate equivalents of another.

2.3 The sample collection

2.3.1 The constraints for sampling

To collect data for this research, sampling method is utilized. In consideration of the time limitation, several constraints are put upon the sample.

Firstly, two bibliographical databases are chosen. One is the National Library of China (NLC for short), which is the biggest public library system in China; the other is the China Academic Library & Information System (CALIS for short), which is the biggest academic library system in China. Both of the library systems use the Chinese Library Classification (CLC) and the Chinese Thesaurus (CT) as indexing tools, which makes it convenient to compare indexing results. The sample in this study is extracted from these two databases.

Secondly, the subject areas of the sample are restricted to botany, computer science, psychology and education. The four subject areas belong to nature science, technology, humanities and social science, respectively. The aim of the restriction is to compare the indexing consistency among different domains.

Subsequently, only monographs are chosen. Serials, periodicals, microform copies, multimedia documents, etc. are excluded. Monographs that are classified and assigned index terms by both NLC and CALIS catalogers are identified. Items with no class numbers or index terms assigned are eliminated.

Finally, only the monographs published from 1988 to 2004 are included. The reason for beginning with 1988 is that since this year computers have been used to catalogue documents in China. That is, since this year there has been an electronic catalogue.

2.3.2 Sampling procedures

Although the constraints for sampling are made, how to draw the pairs of records from the two bibliographical databases is still not clear. At the beginning, it was thought to use keywords “plant”, “computer”, ”psychology”, and “education” to search one catalogue, use the resulting sets to search the other catalogue, and then sequentially pick out the records that measure up to the conditions. But this procedure can make the topics of the sample too general, and the records with specific topics can not be extracted from the databases. As a result, the sample of the study may have a big bias.

In order to avoid the bias of this kind, an alternative method can be adopted as follows: initially, using class numbers to search one catalogue to get the resulting sets, and then using the resulting sets to search the other one. But there are also some problems. First of all, CALIS does not have the classification search function. Simultaneously, although NLC has this function, it is very difficult to use all of the notations to search the database, because the Chinese Classification falls between the two extremes of enumerative and faceted classification. Most important subjects have ready-made class numbers that are enumerated in the CLC schedules. Nevertheless, many subjects are still not provided for, and these can be synthesized by the number-building process. Almost every number in the schedules can be further extended by notation either from one or more of the auxiliary tables or from the schedules themselves. It means that it is very difficult to exhaust the notations to search the database.

Finally, it is decided to use only ready-made class numbers to retrieve. The random sampling method is not used in this research due to the above mentioned constraints. In order to decrease the inherent bias of purposeful sampling, the titles are chosen randomly, as using the resulting sets of the first catalogue to search the other catalogue. Thus, the sample can more possibly reflect the real situation of the whole population.

2.3.3 Size of the sample

With regard to the sample size of this study, in the beginning, it was planned to draw 50 monographs from every year (from 1988 to 2004), and they had to distribute equitably in the four subject areas (botany, computer science, psychology and education). In other words, 850 monographs would be drawn from every subject area. The total number of monographs drawn from the databases would be 3400. But in the end, the total number is 3307, because some records with incomplete subject information are eliminated. Another reason is that too few monographs of the subject area botany are available in NLC. Table 1 presents the distribution of the samples among the four subject areas.

Table 1: Distribution of the samples among the four subject areas

Subject area	botany	computer science	psychology	education
Sample size	759	850	848	850

2.4 Measuring consistency

2.4.1 Consistency measures used in previous studies

In earlier studies, many different formulae to measure indexing consistency have been used. They can be divided into the following two categories:

Pair consistency

When a comparison is made between the indexing of two indexers or between the indexing conducted by the same indexer at different times, various formulae can be used to calculate the pair consistency score. Rolling (1981) enumerated the following 6 formulae:

(1) $\frac{c}{a+b}$

(2) $\frac{c}{a+b-c}$

(3) $\frac{c}{a+b-2c}$

(4) $\frac{2c}{a+b}$

(5) $\frac{2c}{a+b-c}$

(6) $\frac{2c}{a+b-2c}$

In these formulae,
a denotes the number of terms assigned by indexer A
b denotes the number of terms assigned by indexer B
c denotes the number of terms used by both A and B
(a+b) denotes the total number of terms assigned by A and B
(a+b-c) denotes the total number of terms used by either A or B or both
(a+b-2c) denotes the number of terms that have been used by either A or B, but not by both.

Among these formulae, the second one and the fourth one have been most often utilized. And the second one is known as Hooper's formula, the fourth one is known as Rolling's formula.

However, Saarti (2002) used the so-called asymmetrical formula in his study. In this formula the indexing of different indexers is compared in pairs and the number of identical terms used by two indexers is divided by the total number of the terms used by one indexer. If K_1 and K_2 are sets that include the terms used by indexers K_1 and K_2 reading the same document, then the consistency of indexer K_1 in comparison to indexer K_2 is:

$$Consistency_{1,2} = \frac{|K_1 \cap K_2|}{|K_1|} = \frac{\text{the number of identical terms used by indexers } K_1 \text{ and } K_2}{\text{the total number of terms used by indexer } K_1}$$

Correspondingly the consistency of indexer K_2 in comparison to indexer K_1 is:

$$Consistency_{2,1} = \frac{|K_1 \cap K_2|}{|K_2|} = \frac{\text{the number of identical terms used by indexers } K_1 \text{ and } K_2}{\text{the total number of terms used by indexer } K_2}$$

In comparisons of the consistency between several indexers, initially one calculates consistency value for each pair of indexers. After that, the average of these values is calculated for each indexer.

Group consistency

Sometimes, more than two indexers are involved in a comparison. Thus, the consistency measure should be extended. For example, Slamecka & Jacoby (1965), and Markey (1984) compared the indexing consistency among three indexers, and they used the following formula:

$$(\%) \text{of matching terms} = \frac{N(ABC)}{N(A)+N(B)+N(C)-N(AC)-N(BC)-N(AB)+N(ABC)}$$

Where *N (A)*, *N (B)* and *N(C)* equal the number of terms used by each of the three indexers (indexer A, B, or C)

N (AB), *N (AC)* and *N (BC)* equal the number of terms matched between the three pairs of indexers

N (ABC) equals the number of terms matched among all three indexers.

Iivonen (1990) compared indexing consistency among ten indexers using the following formula:

$$X = 100 * \frac{10t + 9/10(t_{abcdefghi} + t_{abcdefghj} + ...) + 8/10(t_{abcdefgh} + ...) + ... + 2/10(t_{ab} + t_{ac} + ...)}{a+b+c+d+e+f+g+h+i+j}$$

Where t= the total number of terms used by all indexers

$t_{abcdefghi}$ = the number of common terms used by indexers A, B, C, D, E, F, G, H and I (no J)

$t_{abcdefghj}$ = the number of common terms used by indexers A, B, C, D, E, F, G, H and J (no I) etc.

$t_{abcdefgh}$ = the number of common terms used by indexers A, B, C, D, E, F, G, H (no I and J) etc.

t_{ab} = the number of common terms used by indexers A, B

t_{ac} = the number of common terms used by indexers A, C etc.

a = the number of terms used by indexer A

b = the number of terms used by indexer B etc.

Although various formulae have been used to calculate group consistency, they can be categorized into two groups. One category is that the formula only calculates the ratio of the number of terms selected by all indexers in the group to the total number of different terms selected for the document (see Slamecka & Jacoby (1965), Markey (1984)). The other category is that the formula takes into account the terms assigned in common by part of the indexers involved in the test, each subset being weighted proportionally to the number of indexers who have been assigning it (see Iivonen (1990)).

2.4.2 Similarity measures in the field of information retrieval

With regard to the vector model of information retrieval, both the documents and queries are conceived as strings of numbers as though they were vectors. Retrieval is based on whether the “query vector” and the “document vector” are “close enough”. Sometimes they are even conceived as sets of terms. A document is retrieved in response to a query, if the document and the query are “similar enough”, i.e. a similarity measure between the document and the query is over some threshold (Dominich, 2001).

For the computation of the similarity between two sets of terms, a number of heuristics have been introduced in literature. These heuristics each have their own way of estimating the degree of overlap between two sets. Typical similarity measures are as follows:

1) Dot product (simple matching coefficient; inner product):

$$SIM_{dotproduct} = |A \cap B|$$

The dot product measure is simply the number of common terms in A and B. It is not normalized.

2) Cosine measure:

$$SIM_{cosine} = \frac{|A \cap B|}{(|A| \cdot |B|)^{1/2}}$$

The cosine measure relates the overlap of the sets A and B to their geometric average.

3) Dice's coefficient:

$$SIM_{dice} = \frac{2 \cdot |A \cap B|}{|A| + |B|}$$

Dice's coefficient relates the overlap of sets A and B to their average arithmetical average.

4) Jaccard's coefficient:

$$SIM_{jaccard} = \frac{|A \cap B|}{|A \cup B|}$$

Jaccard's coefficient expresses the degree of overlap between two sets A and B as the proportion of the overlap from the whole.

5) Overlap coefficient:

$$SIM_{overlap} = \frac{|A \cap B|}{\min(|A|, |B|)}$$

The idea of the overlap coefficient is to determine the degree in which the sets A and B overlap each other.

6) Inclusion coefficient:

$$SIM_{inclusion} = \frac{|A \cap B|}{|A|}$$

The inclusion measure quantifies the degree in which set A is covered by set B. In other words, the measure indicates how good A is a subset of B. It is not symmetric.

Measuring consistency in indexing can be conceived as measuring the similarity of two sets of terms. Thus, these similarity measures in the field of information retrieval can be used for measuring consistency as well. Actually, most of the consistency measures we have mentioned in the subsection 2.4.1 are derived from these typical similarity measures. Among these measures, Dice's coefficient and Jaccard's coefficient have been most widely used in consistency studies. However, in consistency studies, Dice's coefficient and Jaccard's coefficient are known as Rolling's formula (1981) and Hooper's formula (1965), respectively. Thereby, in the following text, we stick to this tradition and call these two formulae Rolling's formula and Hooper's formula.

2.4.3 The two formulae used in this study

In this study, the consistency between the two catalogues is calculated; therefore the formulae of pair consistency are appropriate. As mentioned above, concerning pair consistency measures, both Hooper's and Rolling's formulae have been extensively used. Here we list them again:

1) $Consistency_{Hooper} = c/(a+b-c)$ (Hooper, 1965)

2) $Consistency_{Rolling} = 2c/(a+b)$ (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. The relationship between the two formulae is as follows:

$$Consistency_{Hooper} = Consistency_{Rolling} / (2 - Consistency_{Rolling})$$

Hooper interpreted the degree of agreement between two indexers as the proportion of the terms they commonly assign (*c*) to the distinct terms they totally assign (*a*+*b*-*c*) to a document, while Rolling expressed indexing consistency as the average value of the work of two (or more) indexers. And the average value can be measured by a comparison of the total number of terms correctly assigned by each (*2c*) with the total number of terms assigned by each of them (*a*+*b*). The underlying assumption of Rolling's formula is that only common terms assigned by two indexers are correct. This assumption is controversial and may be problematic. Lancaster (1998) insisted that "quality and consistency are not the same: one can consistently bad as well as consistently good!!" However, in the interest of comparing with previous studies, both formulae are used for calculating indexing consistency in this study.

2.5 Some rules for data analysis

2.5.1 The unit of comparison

How can the indexing results of two indexers be compared? Different methods and various kinds of comparison units have been used in earlier studies. For example, when the indexing language that is involved in a consistency research is a subject heading system, the comparison unit is usually a subject heading string, which can be a main heading (e.g., "Information service") or a combination of a main heading and one or more subheadings (e.g., "Information service- History- Germany"), etc.. When the involved indexing language is a thesaurus, the comparison unit is usually a term, which can be a single word (e.g., "History") or a phrase (e.g., "Information service"). The results are very different when different kinds of comparison units are used. The difference can be shown by the following example:

A book entitled "Plant Tissue Culture" is indexed by two indexers. Indexer A assigns two terms "Plant- Tissue cultures" to it, whereas indexer B assigns four terms "Tissue cultures- Plant- Technology- Theory" to it.

If the comparison unit is a term, according to Hooper's formula, the consistency percentage is 100*2/(2+4-2)=50%; while according to Rolling's formula, it is 100*2*2/(2+4)=66.67%. But if the comparison unit is a combination of terms, then "Plant- Tissue cultures" is totally different from "Tissue cultures- Plant- Technology- Theory". Thus, according to both of the formulae, the consistency percentage is 0%.

That is to say, the indexers are more likely to be consistent in choosing a term than constructing a combination of terms.

Before the indexing results are compared, the comparison unit should be determined. As discussed above, it can be a term or a term string. A term is “a word, phrase, or symbol, especially one used to represent, in a dictionary, catalog, index, or database, a subject or other feature of a work (cited from ODLIS)[1]”. It can be a main heading, a subheading, a descriptor, an identifier, or a keyword, etc.. For instance, “plant”, “technology”, “tissue cultures” are all terms. A term string consists of more than two terms which are combined to represent a specific subject. The combinations are usually according to some syntax rules. An example of a constructed string is a subject heading string that consists of a main heading and one or more subheadings. For instance, “Elementary education- Germany- Bibliography” is a term string.

Obviously, the method of indexing term strings is pre-coordinative indexing. The combination of terms is done by indexers rather than by searchers. “It is typically applied in printed indexes in contrast to electronic retrieval based on Boolean technique, where post-coordinative indexing is normally implemented.” (Olson, 2001) In post-coordinative indexing, the combination of terms is not made during the indexing of the document but during the searching in the database.

Nowadays, most online catalogues support Boolean retrieval. Users are more likely to use the field “keywords” or “descriptors”, rather than “subject headings” to search the database. Therefore, in this research, term is chosen as the comparison unit. On the basis of using a term as the comparison unit, some rules should be explained.

First, duplicates of a term, which is used in different term strings of the same record, are calculated only once. For example, a book entitled “Survey of Contemporary Japanese Education” is assigned the following terms by the indexer of NLC:

Topical subject 1: Education-Survey- Japan

Topical subject 2: Education

---

[1] ODLIS: Online Dictionary for Library and Information Science

The indexer of CALIS assigns only one term "education" to the same book. For the record of NLC, three totally distinct terms are assigned to this book. Although the term "education" is repeated, only one manifestation of it is assessed. Thus, according to Hooper's formula, the consistency between the two indexers in terms of this book is 100*1/(3+1-1)=33.33%; according to Rolling's formula, it is 100*2*1/(3+1)=50%.

Second, the synthetic syntax of terms in a term string is ignored, i.e., how terms are combined is ignored. Meanwhile, in what order terms are combined is also ignored. For instance:

Title of a book: From Diapers to Dating: Beyond the Big Talker, Every Parent's Guide to Raising Sexually Healthy Teens.

Indexer of NLC assigns terms:

Topical subject 1: Sex education- Domestic education- America

Topical subject 2: Sex education

Topical subject 3: Domestic education

Indexer of CALIS assigns terms:

Domestic education, Sex education, America

In this example, the two indexers assign the same three terms to this book. But the order of the three terms is slightly different. The indexer of NLC places "sex education" on the first position, probably because of considering this term as the most important term for this book. But the indexer of CALIS places "domestic education" before "sex education". In view of using a term as the comparison unit, the order of terms and the relationship among terms do not play any important role on the calculation of indexing consistency. So, for this records pair, the consistency is 100%.

Third, only the terms in the field "descriptors" are considered. The indexing policies of both of the library systems permit the use of some enrichment terms that are not in the Chinese Thesaurus. In principle, indexers should place these terms in the field "uncontrolled terms" rather than in the field "descriptors". However, when indexers consider that a term is very important to represent the document in hand, they can place this term in the field "descriptors" as well. In such a case, the unauthorized terms are included in the calculations.

2.5.2 The rules for an exact match

One of the reasons for the variability of the results in previous studies is the variability of the rules for an exact match. Under what kinds of conditions can two terms be counted as an exact match? Owning to different understandings of indexing consistency, the researchers have applied different criterions. Indexing consistency is understood in two different manners, one is terminology consistency, and the other is concept consistency. The consistency on the basis of concepts is usually higher than that on the basis of terms. The reason is that different terms can be used to refer to the same concept, even when the controlled vocabulary is used.

In terms of terminology consistency, the criterions are stringent. Synonyms, abbreviations, spelling variants, and words with different endings or punctuation are regarded as non-matches. Only when two terms match letter for letter, word for word, it is counted as an exact match. However, some researchers have adopted relatively less stringent criterions. For instance, in Tonta's study (1991), variants in spelling (i.e., catalog & catalogue) and punctuation (i.e., on-line & online) were considered as "exact matches", but synonyms were not.

When the comparison unit is a subject heading, the situation becomes a little more complicated. With regard to subject headings, not only the exact matches, but also the partial matches were investigated in former studies. Commonly, when two subject headings have the same main heading, but different subdivision (s), they are treated as "partial matches". Chan (1989) defines "completely matched headings" as a pair of headings that are matched character for character, including field tags, subfield codes, punctuation, and capitalization. She defines "partially matched headings" as a pair of headings that contain identical main headings but different subdivisions, as well as a pair of headings that contain one or more, but not all, identical words in the main heading portion.

With regard to concept consistency, the criterions are less stringent. Iivonen (1990) lists the cases, in which different terms are understood as the same concepts:

(1) if one term is a grammatical or syntactical variation of another,
(2) if one term is a narrower term of another term,

(3) if terms are synonyms.

Detailed information about concept consistency can be obtained from Iivonen (1990).

With respect to the current study, only terminology consistency is investigated. The criterions for an exact match are defined as follows:

Rules for index terms

1) Two terms should match character for character. For example, “relation database” and “relational database” do not match.

2) A term and its abbreviation are not considered as an exact match. For example, “Beijing” is not equal to “Beijing city”.

3) A term with a qualifier does not equal the term itself. For example, “analysis” is not equal to “analysis (chemical)”

4) The clerical errors are ignored.

5) Terms in a separate field “uncontrolled terms” are not taken into account. In both of the library systems, the indexing policies allow for uncontrolled indexing when there are no appropriated terms in the Chinese Thesaurus. These uncontrolled terms should be catalogued in the field “uncontrolled terms”.

The following example shows how these rules are applied to the calculation of consistency:

Title of a book: Mastering Dreamweaver

The terms assigned by the indexer of NLC are as follows:

Topical subject 1: Webpage- Construction- Software tool

Topical subject 2: Webpage

Topical subject 3: Software tool

Uncontrolled term: Dreamweaver

The terms assigned by the indexer of CALIS are as follows:

Webpage, Construction, Application software

In terms of the above mentioned rules, according to Hooper's formula, the consistency is 100*2/(3+3-2)= 50%; according to Rolling's formula, it is 100*2*2/(3+3)= 66.67%. In this example, the uncontrolled term "dreamweaver" is not counted when the consistency is calculated.

Rules for class numbers

1, If the common auxiliary tables' numbers are not consistent, even though the main notations are consistent, they are considered as inconsistent.

For example:

Title of a book: Contemporary Education Philosophy

The indexer in NLC assigns a class number to this book: G40

The indexer in CALIS assigns a class number to the same book: G40-02

In the CLC, "G40" stands for "education theory". "-02" is a number from the common auxiliary table and stands for "philosophy principle".

In terms of this rule, the consistency for class numbers of this pair of records is 0%.

2, Even though the specificity of two notations differs on only one level, they are considered as inconsistent. For example:

Title of a book: Sociology of Adult Education and Continuing Education

The indexer in NLC assigns a class number to this book: G72

The indexer in CALIS assigns a class number to the same book: G720

In the CLC, "G72" stands for "Adult education and extracurricular education", while "G720" stands for "Theory of adult education and extracurricular education".

In terms of this rule, the consistency for class numbers of this pair of records is 0%.

3, The punctuation errors and clerical errors are ignored.

For ease of reading and transcription, the class numbers are usually punctuated after every third digit, e.g. TU236.347.8. Some times an error occurs, when the class number is very long. In the sample of this research, it is found that some class numbers are only punctuated after the first third digit. This kind of punctuation errors are ignored in this study.

2.6 Statistical analysis

After the data are collected and the consistency is calculated, they will then be subjected to statistical analysis. On the one hand, descriptive statistics will be conducted, i.e. the data will be summarized either numerically (e.g. mean and standard deviation) or graphically (e.g. various kinds of charts and graphs). On the other hand, inferential statistics, for instance, hypothesis testing, correlation, regression, ANOVA, and so on, will be conducted as well.

In this research, many variables are involved, among which the variable "consistency" is a dependent variable, the others are all independent variables. Concerning the t-tests conducted in this study, only when the probability is less than the preset alpha level (usually .05), the results are considered statistically significant.

Moreover, Pearson correlation coefficient (r) is adopted to reveal whether two variables tend to vary together. And according to Cohen (1988), when r is larger than 0.70, the strength of the correlation is very large; when r is less than 0.70, but larger than 0.50, the strength is large; when r is between 0.30 and 0.50, the strength is medium, and when r is between 0.10 and 0.30, the strength is small.

2.7 Conceptual analysis

Sometimes we can only make some inferential conclusions from the statistical results. We still cannot say what the exact reasons for the inconsistencies are. In order to make it clearer, the sample is analyzed record by record. The terms and the notations assigned are analyzed conceptually, so that the direct reasons for the inconsistencies can be found out. Some examples with detailed comparisons of terms are given to help reader better understand this issue.

3 Results

3.1 Introduction

Firstly, the overall consistency of index terms and that of notations is described in Section 3.2. From the descriptive statistical results we can get the overview of this research problem. The consistency of this study is a little higher than the results from earlier studies. Then, in Section 3.3 some of the factors that may affect indexing consistency are analyzed. And how these factors influence indexing consistency is explored. The results show that in earlier studies some statements about the relationships between indexing consistency and the factors are problematic, because sometimes the relationship between indexing consistency and a factor is not direct or simple, but very complicated. In the end, different categories of indexing inconsistencies are closely examined in Section 3.4, so that the causes that directly result in inconsistencies can be discovered.

3.2 Descriptive statistical results

3.2.1 Consistency of index terms

Consistency of index terms between the indexers of NLC and the indexers of CALIS with regard to the 3,307 monographs is examined and measured 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% (See Table 2).

Table 2: Descriptive statistics of the overall consistency of index terms (according to two formulae)

	according to Hooper's formula	according to Rolling's formula
No. of books	3,307	3,307
Mean	64.21	70.71
Std. Error of Mean	0.64	0.59
Std. Deviation	37.04	33.86
Skewness	-.41	-.87
Kurtosis	-1.35	-.49
Median	66.67	80.00
Mode	100.00	100.00
Minimum	.00	.00
Maximum	100.00	100.00

The standard deviation, which is the most common measure of statistical dispersion, is also examined. According to Hooper's formula and Rolling's formula, they are

37.04% and 33.86%, respectively. The two values are very similar and relatively high. It indicates that the consistency scores of index terms are relatively far from the mean, i.e, they are highly variable (See Figure 1 and Figure 2).

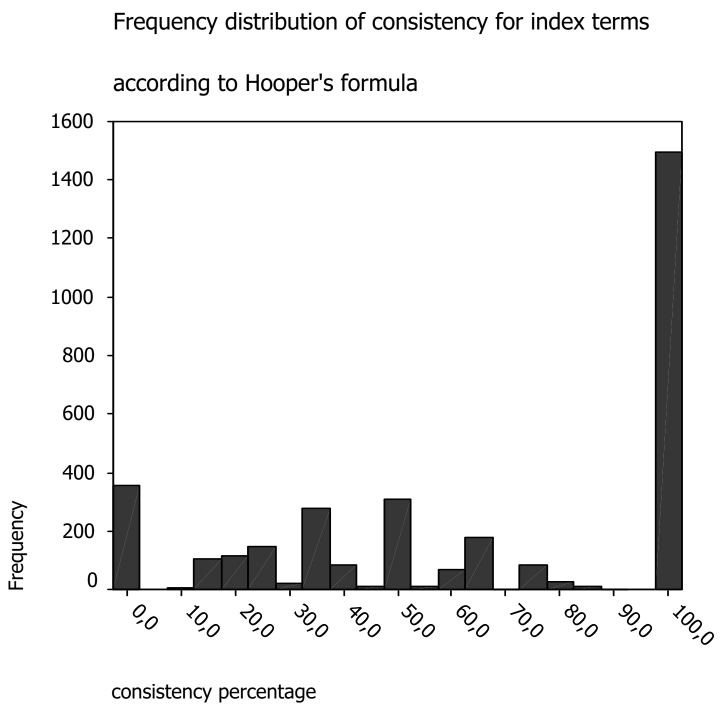

Figure 1: Frequency distribution of consistency for index terms (according to Hooper's formula)

In addition, another important descriptive statistic, i.e. skewness, is conducted. It is important to know if the variables are highly skewed, because most common inferential statistics (e.g., t-test) assume that the dependent variable is normally distributed. The skewness of the distribution of consistency in index terms is negative (-0.41 and -0.87 according to Hooper's formula and Rolling's formula, respectively), which is shown by the long lower tail in Figure 1 and Figure 2. It indicates that in most cases the consistency is low.

Moreover, the kurtosis is calculated. Although the values of kurtosis are different when the consistency is calculated according to different formulae, they are negative in both cases (-1.35, -0.49). Both Figure 1 and Figure 2 demonstrate a distribution with smaller "peak" around the mean and "thin tails", which indicates that the consistency values are more likely to distribute in the tails of distribution.

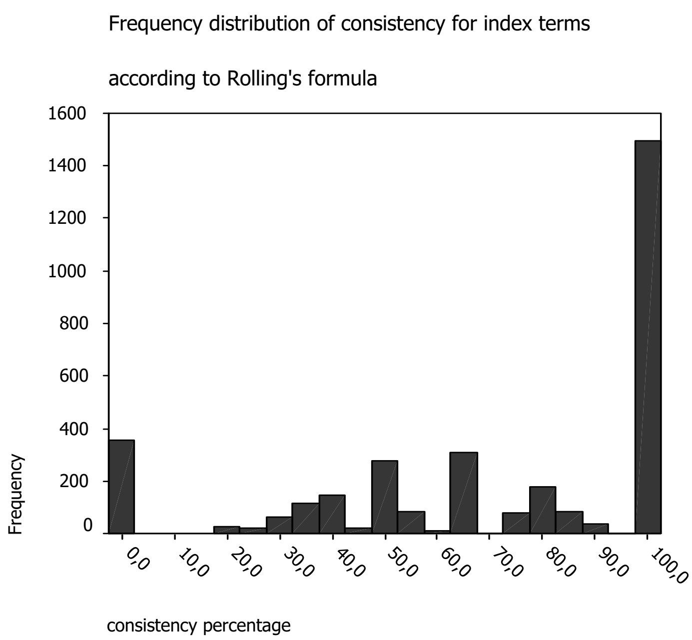

Figure 2: Frequency distribution of consistency for index terms (according to Rolling's formula)

Although both the skewness and the kurtosis are negative, they are still within the -2 to +2 range, i.e. the data can still be conceived as normally distributed.[2] Therefore the assumption of normality is not violated. However, from Figure 1 and Figure 2 we can see that in many cases the consistency is 0% or 100%. Hence, the records with 0% or 100% consistency percentage are divided from the whole sample and discussed separately. So the whole sample is divided into three parts. Part A consists of the records with 0% consistency. Part B consists of the records with 100% consistency. The rest records (i.e., where consistency is larger than 0% and smaller than 100%) belong to Part C. The descriptive statistics of Part C see Table 3.

Table 3: Descriptive statistics of overall consistency for index terms (where consistency of index terms is larger than 0% and smaller than 100%)

	according to Hooper's formula	according to Rolling's formula
No. of books	1,455	1,455
Mean	43.19	57.96
Std. Error of Mean	0.49	0.48
Std. Deviation	18.67	18.3
Skewness	.28	-.09

[2] http://www2.chass.ncsu.edu/garson/pa765/assumpt.htm

Kurtosis	-.98	-1.05
Median	40.00	54.17
Mode	50.00	66.67
Minimum	10.00	18.18
Maximum	88.89	94.12

Figure 3 and Figure 4 show the frequency distribution of consistency (where consistency is larger than 0% and smaller than 100%) for index terms according to the two formulae. Markedly, the distribution is a normal distribution, although it is slightly skewed and flatted. Since the records with 0% and 100% consistency are not included, the mean consistency of this part is notably lower than the overall consistency. And the standard deviation is also smaller.

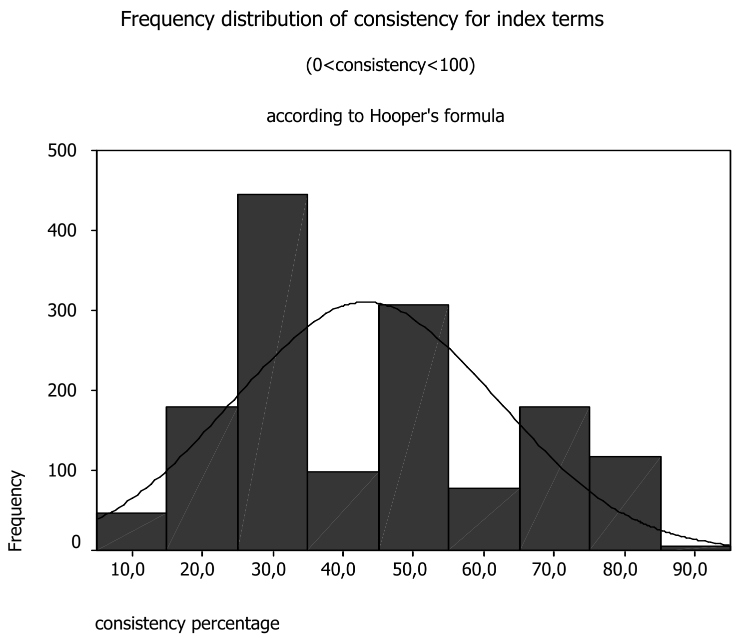

Figure 3: Frequency distribution of consistency for index terms (where consistency of index terms is larger than 0% and smaller than 100%, according to Hooper's formula)

Besides, it is notable that almost half of the whole sample (1,495 out of 3,307, i.e. 45.21%) has 100% consistency, while 10.8% of the whole sample (357 out of 3,307) has 0% consistency. Both of the parts are worthy of investigating, so that the reasons for consistency and inconsistency can be discovered.

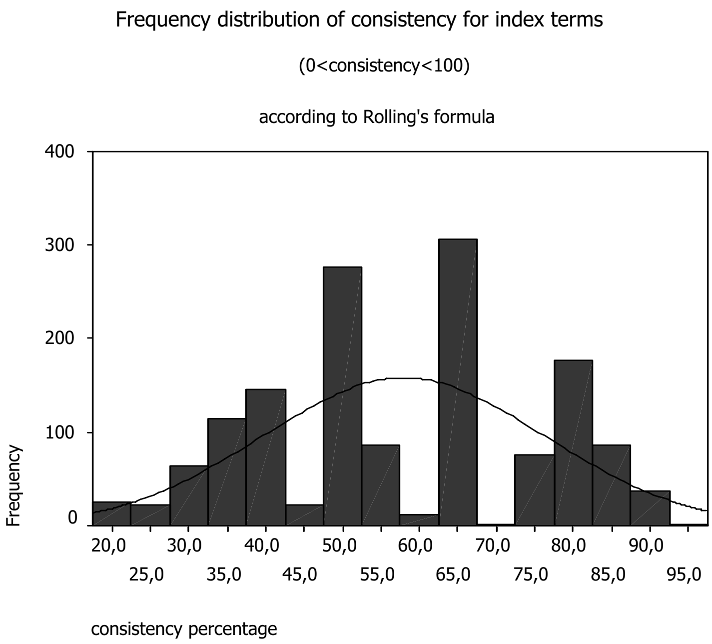

Figure 4: Frequency distribution of consistency for index terms (where consistency of index terms is larger than 0% and smaller than 100%, according to Rolling's formula)

3.2.2 Consistency of notations

Except for the consistency of index terms, the consistency of notations between the indexers of NLC and the indexers of CALIS with regard to the 3,307 monographs is also examined and measured with the use of the two formulae described above. According to Hooper's formula, the overall consistency of notations is 61.58%, while according to Rolling's formula it is 63.39% (See Table 4).

Table 4: Descriptive statistics of consistency for notations (according to two formulae)

	according to Hooper's formula	according to Rolling's formula
No. of books	3,307	3,307
Mean	61.58	63.39
Std. Error of Mean	0.80	0.79
Std. Deviation	45.82	45.55
Skewness	-.47	-.59
Kurtosis	-1.65	-1.54
Median	100.00	100.00
Mode	100.00	100.00
Minimum	0.00	.00
Maximum	100.00	100.00

When Hooper’s formula is used to calculate consistency, the standard deviation is 45.82%, whereas it is 45.55% when Rolling’s formula is used. The two values are almost the same, and are very high, even higher than the standard deviation of consistency for index terms. The results show a bipolar distribution.

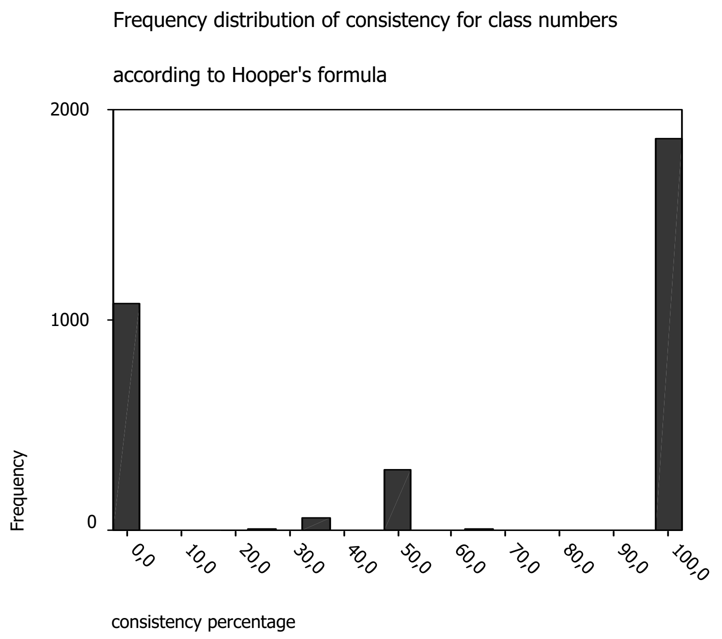

Figure 5: Frequency distribution of consistency for class numbers (according to Hooper’s formula)

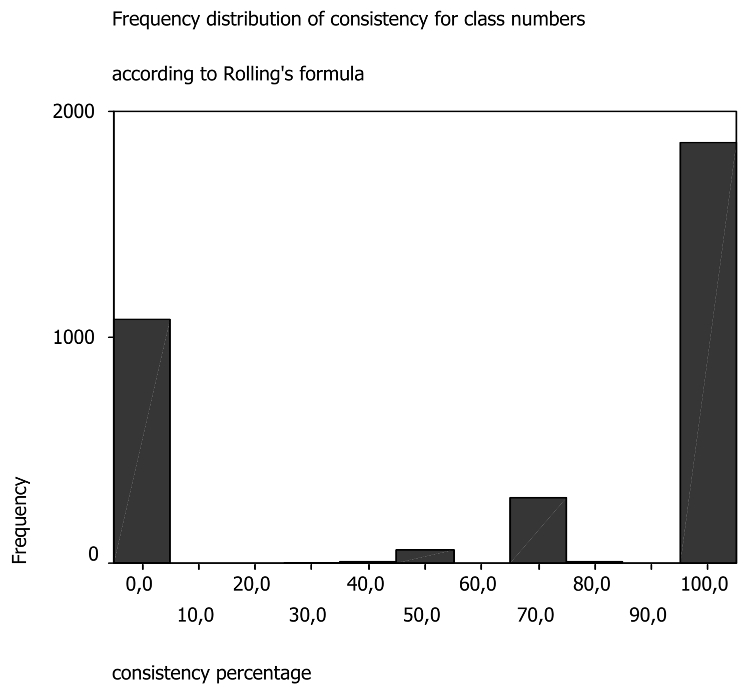

Figure 6: Frequency distribution of consistency for class numbers (according to Rolling’s formula)

From Figure 5 and Figure 6 we can see that 1,080 (out of 3,307, i.e. 32.66%) pairs of records have 0% consistency and 1,865 (out of 3,307, i.e. 56.4%) pairs have 100% consistency. Only 362 (out of 3,307, i.e. 10.95%) pairs of records have the consistency value between 0% and 100%. The reason is that in most cases only one class number is assigned to a book. When the two indexers are consistent, the consistency is 100%; when they are not consistent, the consistency is 0%.

3.3 Factors influencing indexing consistency

3.3.1 Introduction

Many factors can affect indexing consistency. This is one of the reasons for the great variability of indexing consistency in previous studies. Zunde and Dexter (1969) discussed the factors influencing indexing consistency in detail. They categorized the variables into three groups: semantic factors, pragmatic factors and environmental factors. The following are the 25 factors they listed:

Semantic factors
– Author's vocabulary
– Readability of text
– Subject area
– Specificity of discourse
– Author's style and clarity of presentation
– Redundancy of information in text
– Indexing method
– Constraints imposed on indexers in their choice of indexing terms
– Indexing aids of semantical nature, such as thesauri, dictionaries, etc.

Pragmatic factors
– General educational background of the indexers
– Maturity (age, etc.)
– Personality
– Personal history

– Aptitude for indexing
– Motivation in work
– Disposition
– Fatigue, health conditions, etc.
– Psychological constraints, interests, etc.
– Significance judgment related to document structure

Environmental factors
– Noise in the work area
– Climatic conditions (temperature, humidity, etc.)
– Lighting conditions
– Interruptions in the indexer's work
– Time constraints imposed on the indexers
– Type of technical equipment, etc.

Although they have listed these factors, they have not discussed all of them. In former consistency studies, only some of the factors have been investigated. Moreover, in the light of how these factors affecting indexing consistency, the researchers in this field are not totally of one opinion. Further research on these factors is needed. Hence, some of these factors are investigated in this study, and the results show that some widely accepted statements in earlier studies are not totally correct or that some of the issues have been oversimplified.

3.3.2 Exhaustivity of indexing

In many cases, a document deals with a number of concepts. How many of the concepts should be represented? It is one of the problems that indexers or cataloguers should face. Exhaustivity of indexing (also called depth, or level of indexing) refers to "the proportion of concepts covered in assigning index terms or the breadth of subject matter covered." (Olson, 2001) Exhaustivity correlates strongly to the number of terms assigned to a document, although it is not equal to the number of terms assigned. Generally, the more terms are assigned, the higher level of exhaustivity is achieved.

There is a trade-off between high and low exhaustivity. Lower exhaustivity can result in lower recall, while higher exhaustivity can result in lower precision. When a concept is not indexed, the documents dealing with this concept can not be retrieved. On the other hand, when a peripheral concept is indexed, the retrieved documents may not be relevant. Jones (1972) suggested that there should be an optimum level of indexing exhaustivity for a given document collection. It can be achieved by adjusting the average number of descriptors per document, so that the chances of requests matching relevant documents are maximized, while too many false drops are avoided.

The effects associated with the number of subject headings have been studied in several different ways. There is some literature addressing the relationship between the number of subject headings and circulation figures. Banks (2004) found that "the optimal number of subject headings on bibliographic records appears to be one or two, which generated the largest percentage of circulation among the books in the tested sample".

Exhaustivity of indexing is a very important issue. It affects not only retrieval effectiveness and circulation, but also indexing consistency. Various investigations have explored the relationship between indexing consistency and exhaustivity. It is widely believed that higher exhaustivity leads to lower indexing consistency. In other words, the more terms indexers assign to a document, the lower indexing consistency they achieve. This sounds plausible and logical. However, the relationship between indexing consistency and exhaustivity is not as straightforward as it might seem. The following paragraphs will show that this issue has been oversimplified.

3.3.2.1 Number of index terms assigned

3.3.2.1.1 Average number of terms assigned

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). Figure 7 displays the frequency distribution of the number of index terms. It is notable that in most cases (65%) only two or three terms are assigned to a book. In comparison with other databases, too few terms are assigned.

Some traditional guidelines for exhaustivity which are followed by the indexers may contribute to the low level of exhaustivity with regard to this whole sample. One of them is the rule of summarization. As Olson (2001) said, "Library catalogs differ drastically from other online databases in relation to exhaustivity. The subject indexing in online catalogs is guided by the notion of summarization rather than exhaustive indexing, that is, assignment of only one subject heading that represents the total content of the bibliographic item is preferred to several subject headings representing parts of the item." Another one is the "rule of three", i.e. up to three specific concepts in the same hierarchy may all be enumerated, but if there are four or more specific concepts in the same hierarchy the next broader concept should be represented.

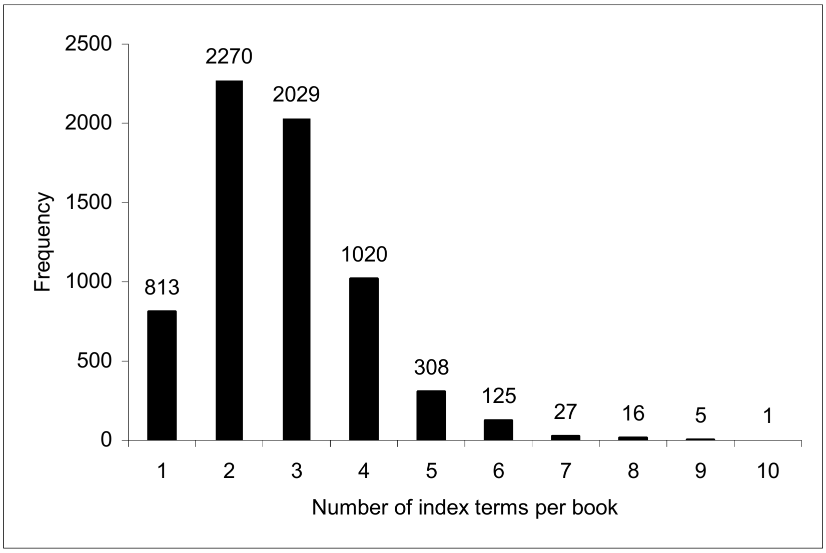

Figure 7: Frequency distribution of the number of index terms per book (the whole sample)

According to the "Guideline for the subject cataloging of Chinese documents", which is used by CALIS, no more than ten terms can be assigned to a document. At the same time, this guideline indicates that summarization of a document's overall content is a basic objective of subject cataloging. It is the main reason for the low level of exhaustivity with regard to this whole sample.

3.3.2.1.2 Comparison of the two catalogs

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 catalogs with regard to the number of index terms. The frequency distribution of the number of terms is shown by Figure 8. The two curves in Figure 8 are very similar, which indicates that the two catalogs have identical level of indexing exhaustivity. It may be due to obeying the same policy of indexing exhaustivity.

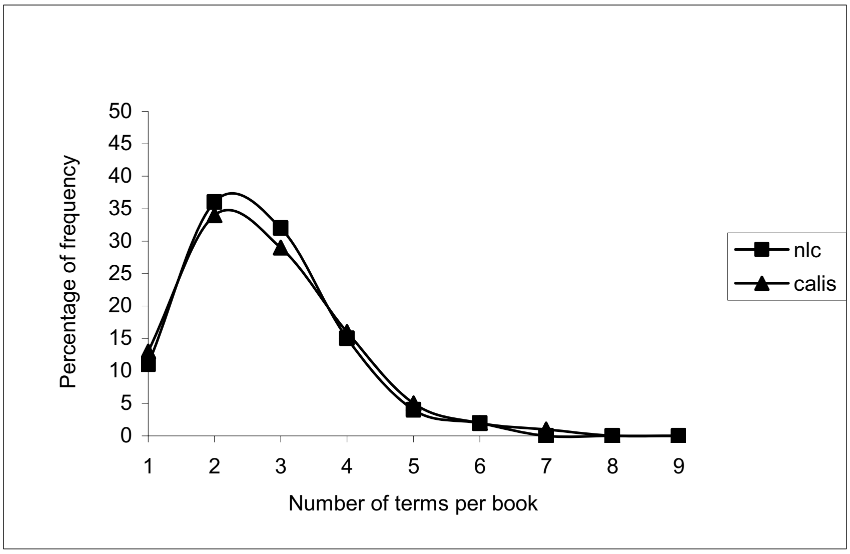

Figure 8: Relative frequency distribution of the number of terms per book (comparison of the two catalogs)

3.3.2.1.3 Comparison of different subject areas

With respect to the four subject areas observed in this study, it is interestingly noticeable that the books in education are indexed most exhaustively, while the books in psychology are indexed least exhaustively (see Table 5). The difference is significant at 0.01 level (t =22.93, df =3,394), which is clearly shown by Figure 9. And there is no big difference between computer science and psychology.

Table 5: Average number of terms assigned per book, according to domains

Domain	Number of books	Average number of indexing terms	Standard deviation	Standard Error
Education	1,700	3.10	1.28	0.03

Botany	1,518	3.06	1.18	0.03
Computer science	1,700	2.66	1.13	0.03
Psychology	1,696	2.21	0.98	0.02

Basically, the terminology in soft sciences is more ambiguous and unstable, and we need more terms to describe a document precisely. Inversely, we need less terms for documents in hard sciences. As Losee (2004) said, "one needs fewer subject-bearing phrases when the phrases contain more precise language, such as one may find in the harder sciences and medicine." Thus, theoretically, we need more terms to describe a book in education and psychology than in botany and computer science.

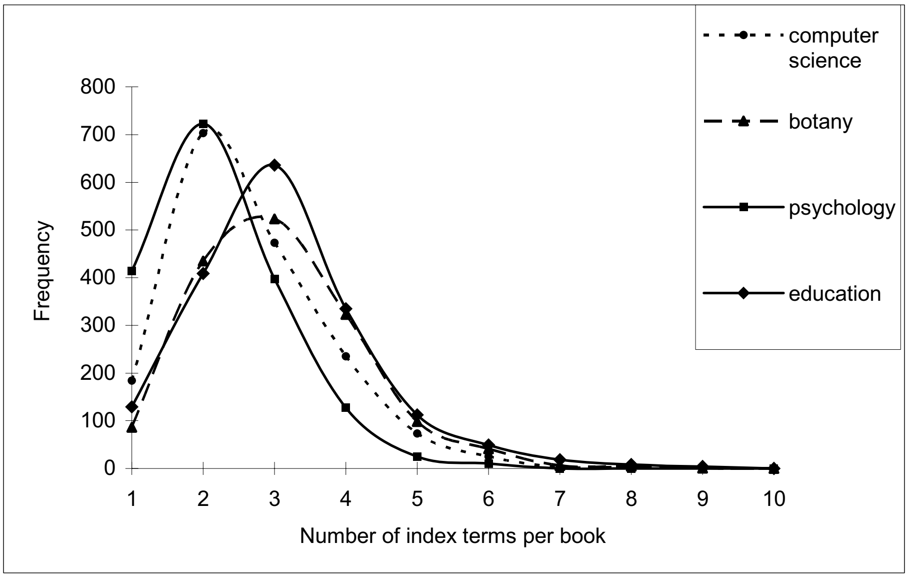

Figure 9: Frequency distribution of the number of index terms (comparison of the four subject areas)

However, the results are not the same as expected. It may be due to the different size of vocabularies of the four subject areas. For instance, there are 1,514 terms available in the vocabulary of botany, whereas there are only 492 terms available in the vocabulary of computer science. Because the complete volumes of the Chinese Thesaurus are not available in Germany, the number of terms in the vocabularies of education and psychology can not be counted. However, from the comparison of botany and computer science, we can speculate that there is some correlation between the number of index terms assigned and the number of terms available from the vocabulary.

3.3.2.1.4 Comparison of different years

From Figure 10 we can see that there is a tendency of an increase in the average number of index terms over the years. The mean value increases from 2.37 terms per record in 1988 to 3.03 terms in 2004. Although the slope of the regression line is very small (namely 0.0422), the adjusted R^2=0.84 is large. According to Cohen (1988) the correlation between the average number of index terms and year of publication is very strong.

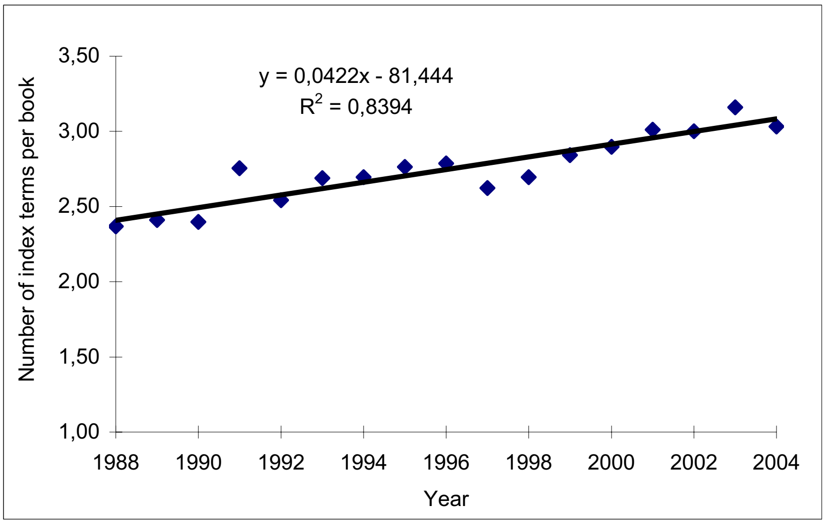

Figure 10: The increase in the average number of index terms per book over the years

The increase of exhaustivity over the years has also been proved by earlier studies. It has been indicated that the average number of subject headings per book in major bibliographic databases (National Union Catalog, OCLC, MARC database, etc.) has increased from 1.3 subject headings per book for the period 1950-1973 to 1.4 by 1980, rising to an average of 2 by 1989, with the latter figure breaking down to an average of 2.14 subject headings per title assigned by LC, and 1.9 subject headings by other libraries. (Shubert, 1992; Drabenstott and Vizine-Goetz, 1994).

There are many reasons for the increase in indexing exhaustivity. First, it may be due to the development of technology. As discussed above, the practice of summarization and the rule of three are the factors that limit the number of index terms. Nevertheless, they stem partly from the days of card catalogs. In the days of card catalog, more subject headings or descriptors increased the labor of filing and occupied the physical space. Nowadays the space limitations of the card catalog are of little or no concern in OPACs. Therefore, there is no reason to limit the depth of analysis or to restrict the number of subject headings assigned to a document. In a word, the possibilities of electronic retrieval have encouraged the librarians to index more deeply and exhaustively than heretofore.

Second, the specificity of subjects in documents has been increasing, more and more interdisciplinary fields have been created, and different viewpoints have been coming up. All these complicating factors can influence the indexing exhaustivity, because one needs more terms to represent the complex subjects than the simple ones.

Third, the indexing policy has a direct impact on indexing exhaustivity. The development of technology and the change of contents in documents require the change of the indexing policy. Olson (2001) described such changes in the Library of Congress as follows: “LC policy moved away from the traditional summarization notion, and the proportion of works with more subject headings increased. LC’s current policy declares a maximum of six subject headings as generally appropriate but permits up to ten subject headings per work, sometimes even more.”

3.3.2.1.5 Relationship between number of index terms and indexing consistency

In earlier studies, it has been proved that as the number of terms decreases, indexing consistency increases (Leonard, 1975). But this may be oversimplified. Lancaster (1998) assumes that if terms are assigned in order of priority, indexing consistency may peak at the level of two terms and then begin a gradual decline up to a point when so many terms have been assigned will again increase.

The data in this study shows whether two indexers assign equal number of index terms to the same book profoundly influences indexing consistency. With reference to Figure 11 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.

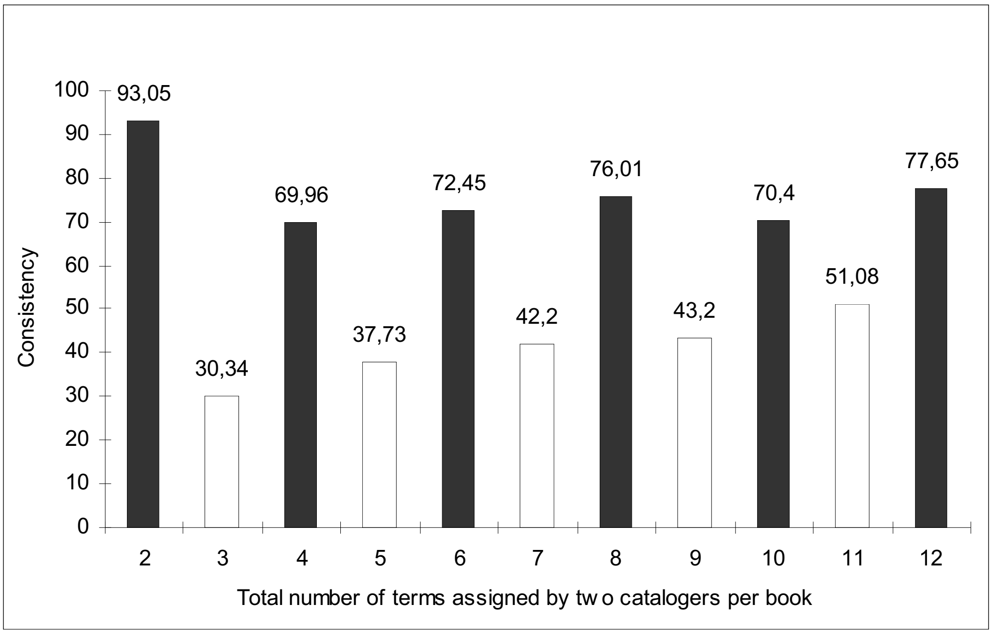

Figure 11: Consistency and total number of terms assigned by two indexers per book

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. This is justified by Table 6.

Table 6: Average consistency and number of terms assigned by two indexers per book

CALIS / NLC	1term	2terms	3terms	4terms	5terms	6terms	7terms	8terms	9terms
1term	93.02	31.51	23.61	7.5	20	8.33			
2terms	29.52	75.47	40.87	28.15	18.27	22.86			
3terms	24.29	36.9	80.14	48.46	35.84	32.86	26.98	10	
4terms	18.75	36.12	42.51	84.15	51.84	32.94	39.25	33.33	
5terms	6.67	24.85	31.58	44.82	80.93	62.42	43.33	37.22	55.56
6terms		23.81	16.43	54.76	55.34	88.24	77.98	55.56	36.36
7terms					50	85.71	100	68.06	60
8terms			16.11		18.18			88.89	
9terms						0			88.89

In this table, the column heading represents the number of terms assigned in the NLC, while the row heading stands for the number of terms assigned in the CALIS. The interception of the row and the column is the mean value of consistency. The backward slashes indicate that there is no entry in the sample under such situation. For instance, the value 93.02 in the interception of the first column and the first row is the mean consistency when the indexers in both library systems assign only one term to a book. Actually all the values in the diagonal cells are the mean values of consistency when two indexers assign equal number of terms to a book. It is notable that these values are the highest ones in the whole table. Meanwhile, the closer the values are to the diagonal cells, the larger they are, although there are some exceptions. It means that when two indexers have identical level of exhaustivity, the indexing consistency is apparently high.

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.

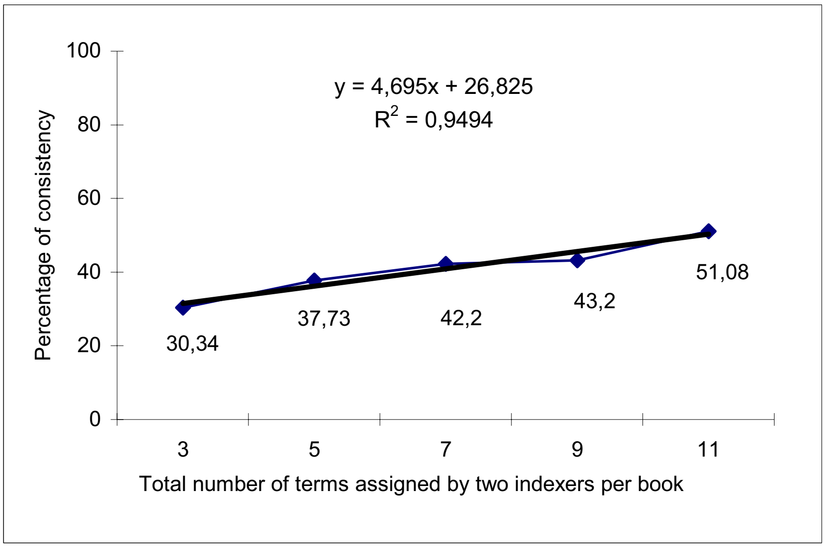

Figure 12: Relationship between consistency and total number of terms assigned by two indexers per book (odd number)

Under condition 1, there is a positive correlation between consistency and exhaustivity, which is illustrated by Figure 12. It indicates that the more terms two indexers totally assign to a book, the more likely they assign the same terms. The explanation for this phenomenon may be that assigning more terms decreases the possibility of getting zero consistency.

Under condition 2, the relationship between consistency and the total number of terms assigned is a little more complicated. It is shown by Figure 13 that 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 10 terms.

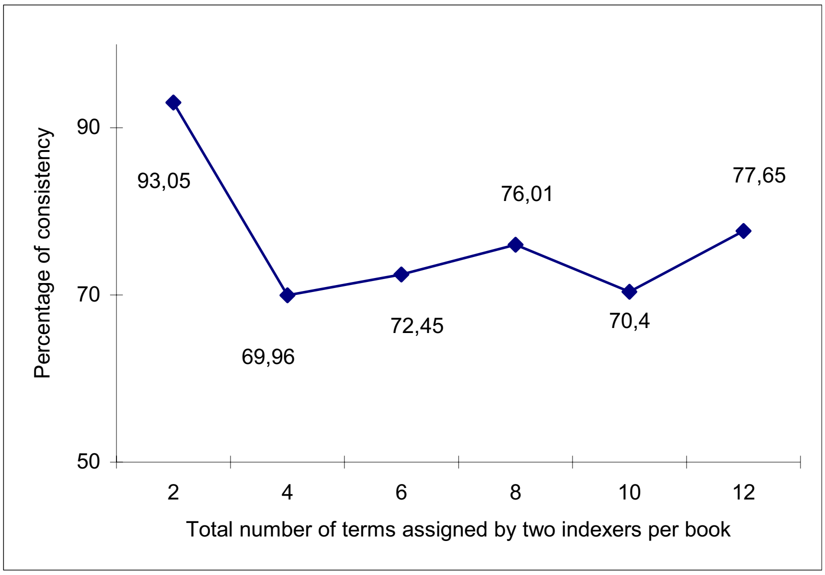

Figure 13: Relationship between consistency and total number of terms assigned by two indexers per book (even number)

From Figure 12 and Figure 13, it can be seen that 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.

3.3.2.1.6 The influence of the formulae used on the relationship between number of index terms and indexing consistency

The experimental data have 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. Actually, it is inevitable with the use of the two formulae that have been introduced in Subsection 2.4.3, 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:

$$E(consistency) = \sum_{c=0}^{10} \frac{1}{11} * \frac{c}{20-c}$$

In this case, the expect 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 have to achieve high consistency. If we plot the experimental data in Table 6 on a graph (see Figure 14), we can see it more clearly.

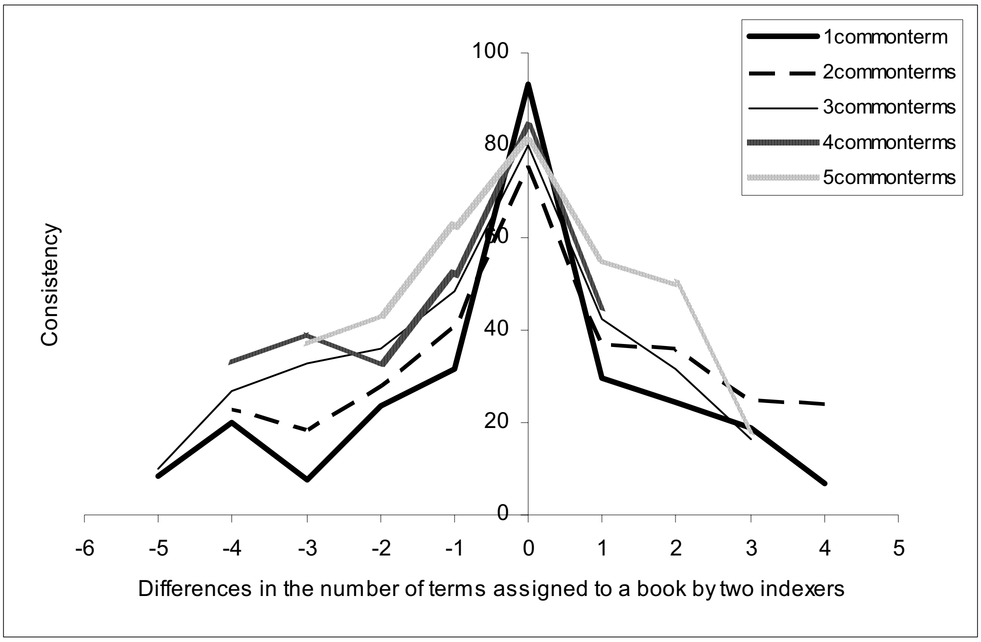

Figure 14: The smaller differences in the number of terms assigned by two indexers, the higher consistency can be achieved

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 15.

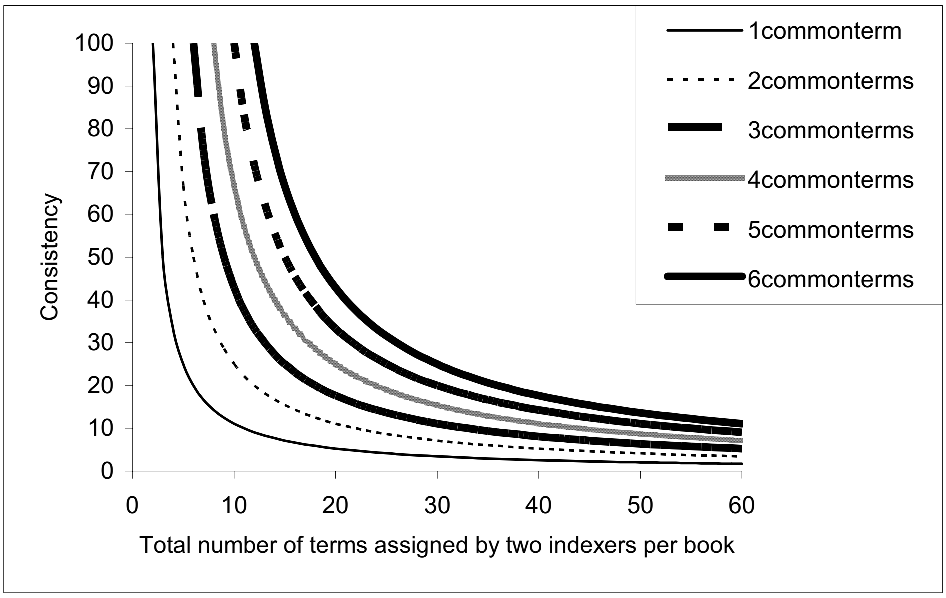

Figure 15: The consistency decreases with the increase of the total number of terms assigned to a book by two indexers, when the number of terms commonly assigned is constant

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 15 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 15 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.

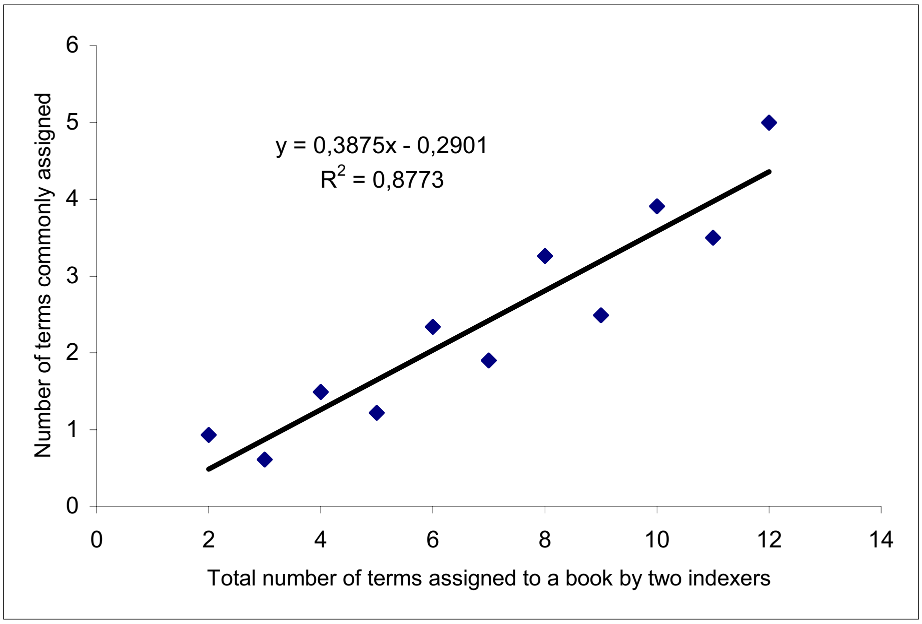

Figure 16: Relationship between the total number of terms assigned and the number of terms commonly assigned per book by two indexers

The experimental data show that the more terms two indexers totally assign to a book, the more terms they commonly assign (see Figure 16). 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, … ,(a+b)/2}, where $\underline{a+b}$ is even; or c= {0, 1, 2, …, (a+b-1)/2}, where ($\underline{a+b}$) is odd. If we calculate the expected value of $\underline{c}$, we can find that the larger $\underline{a+b}$ is, the larger the expected value of $\underline{c}$ is.

The experimental data also show that the more common terms two indexers assign to a book, the more consistent they are (see Figure 17). 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.

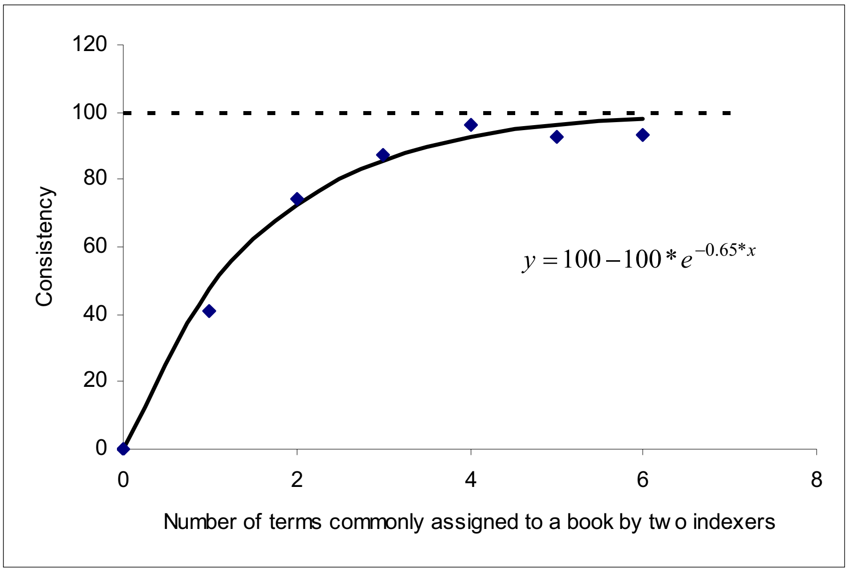

Figure 17: Relationship between the number of terms commonly assigned to a book by two indexers and indexing consistency

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.

To sum up, the formulae used in indexing consistency studies profoundly influence the relationship between indexing exhaustivity and consistency. It is unavoidable that two indexers are more likely to be consistent with each other when they have identical level of exhaustivity. Moreover, the smaller differences are between them considering the number of terms they assign to a document, the greater probability they have to achieve high consistency. In addition, the more terms they totally assign to a document, the smaller are the differences in the consistency scores, no matter how many terms they commonly assign.

3.3.2.2 Number of notations assigned

A total of 7,600 notations are assigned to the 6,614 records, i.e. 3,307 books, averaging 1.15 notations per record. The indexers of NLC assign 3,712 notations, averaging 1.12 notations per book, while the indexers of CALIS assign 3,888 notations, averaging 1.18 notations per book. The level of indexing exhaustivity of NLC is considerably lower than that of CALIS concerning the average number of notations assigned.

Figure 18 shows the frequency distribution of number of notations. It is notable that in most cases (85.68%) only one class number is assigned to a book. It is a result of classification convention, namely, the original purpose of library classification is to arrange books on shelves. Since a book can only have one position, only one class number can be assigned to a book.

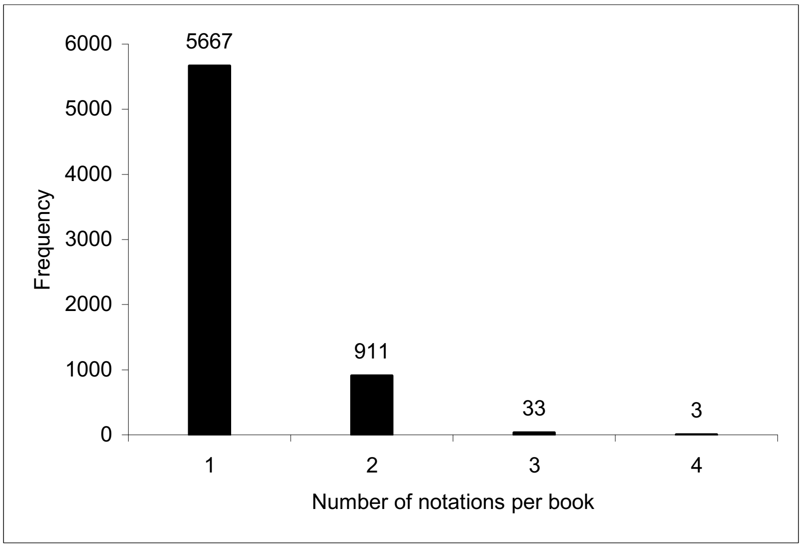

Figure 18: Frequency distribution of the number of notations per book

However, as the development of information retrieval technology and interfaces of OPACs, there have been some possibilities of using library classification to retrieve machine-readable bibliographic records. And a difference has been made between the bibliographic retrieval notation (class number) and the shelving notation (call number), so that more than one retrieval notation can be assigned to represent the compound topics. To some extent it can resolve the problem of linearity of enumerative classification schemes. Accordingly, libraries encourage indexers to assign more notations to a document. By reference to Figure 19 we can see that the average number of notations is increasing over the years. The mean value increases from 1.07 notations per record in 1988 to 1.18 notations in 2004. But it peaks at 1.22 in 1999 and 2000. Whereas it permits more than one bibliographic retrieval notation to a document in OPACs, indexers are inclined to assigning one notation.

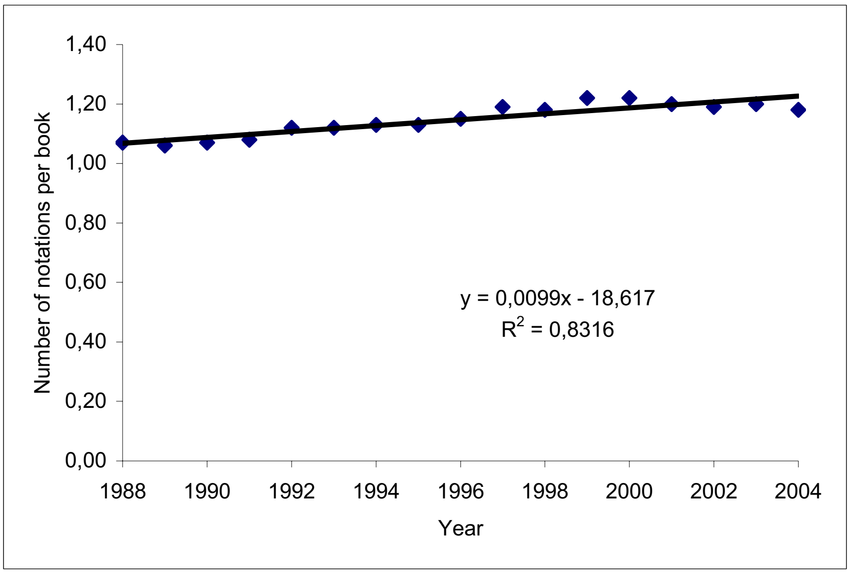

Figure 19: The increase in the average number of notations per book over the years

The relationship between number of notations and indexing consistency is studied as well. Inspection of Figure 20 indicates that when the number of notations assigned to a document by two indexers is equal, the indexing consistency is substantially high. On the contrary, when the number of notations is unequal, the consistency is significantly low. This result is the same as that of index terms. At the same time, we can see from Figure 20 that the more notations two indexers totally assign to a book, the higher consistency is achieved.

According to the discussions about the relationship between indexing consistency and the number of index terms and notations, we can summarize that higher exhaustivity brings about higher indexing consistency, rather than leads to lower indexing consistency.

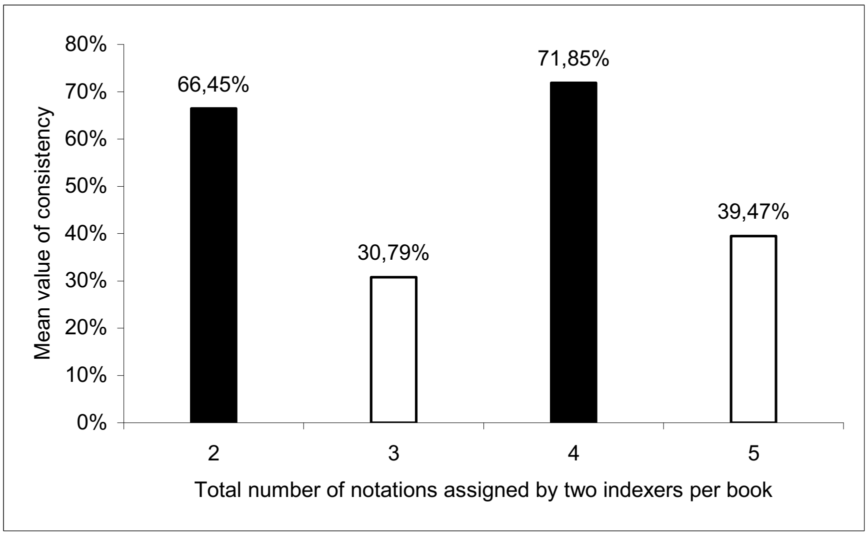

Figure 20: Relationship between consistency and total number of notations assigned by two indexers per book

3.3.3 Specificity

The definition of specificity has been debated for a long while. Marshall (2003) summarized five different definitions of specificity. Actually these definitions are defined from different viewpoints.

The first definition of specificity concerns with type of entry, namely direct entry or indirect entry. It is very known that Charles Cutter stated in his rules for a dictionary catalog that specific entry should be used when assigning subject headings. But there is a misunderstanding. “Specific entry” here is referring to “direct entry”, rather than specific. An indirect entry can be as specific as a direct entry.

The second definition of specificity is relative to the content of the document being cataloged, in fact, it is coextensivity. Coextensivity is the extent to which term(s) assigned express the content exactly. This kind of specificity is in relation to the document being cataloged. It is the relationship between terms assigned and the document to which they are applied.

The third definition of specificity is relative to the indexing vocabulary. A term has an inherent level of specificity concerning its position in a thesaurus. The lower level in a hierarchical chain a term is on, the more specific it is. It is a semantic property of index terms.

The fourth definition of specificity is in relation to frequency of application in indexing. It was initially advanced by Jones (1972). She defined that "the specificity of a term is the number of documents to which it pertains." It is a statistical property of index terms.

The fifth definition of specificity is in relation to the demands of an information storage system. It was advocated by Angell (1954). He called this kind of specificity "the specificity that is desirable".

We discuss specificity in the light of the fourth definition, which is called "statistical specificity".

3.3.3.1 Statistical specificity

Initially, specificity is an issue to be considered when vocabularies are constructed, that is, an optimum level of specificity should be determined when a vocabulary is set up so that a collection of documents can be described precisely with this vocabulary. In this sense, specificity is a semantic property of index terms.

Gradually, some researchers begin to realize that "index term specificity must, however, be looked at from another point of view." Jones (1972) redefined the specificity of a term as the number of documents to which it pertains. She said: "A frequently used term thus functions in retrieval as a non-specific term, even though its meaning may be quite specific in the ordinary sense." Obviously, Jones' new definition of specificity places greater emphasis on the value of a term as a means of distinguishing one document from another than on its value as an indication of the content of the document itself (Jones, 1972). It is a statistical property of index terms. According to Jones, more documents are gathered under the general terms than under the specific terms, that is to say, the less specific terms have a larger collection distribution than the more specific ones. However, when terms are very heavily used, they are not discriminating any more, i.e. they become less effective as a means of retrieval.

Luhn (1958) also noticed that the statistical specificity of terms could affect retrieval. He suggested that after ranking by frequency of occurrence all the words contained in a collection or a sample of a collection, it was possible to establish two cutoffs: one at the upper region, where words are so common that they are useless in distinguishing document sets, and one at the lower region, where words are so uncommon that they do not serve to cluster documents in a useful way. He referred to this characteristic of words as their "resolving power" or their "degree of discrimination". Figure 21 shows this situation, where line C is the high-frequency boundary and line D is the low-frequency boundary, and where the words whose frequencies fall in the middle range, E, are considered to have acceptable resolving power.

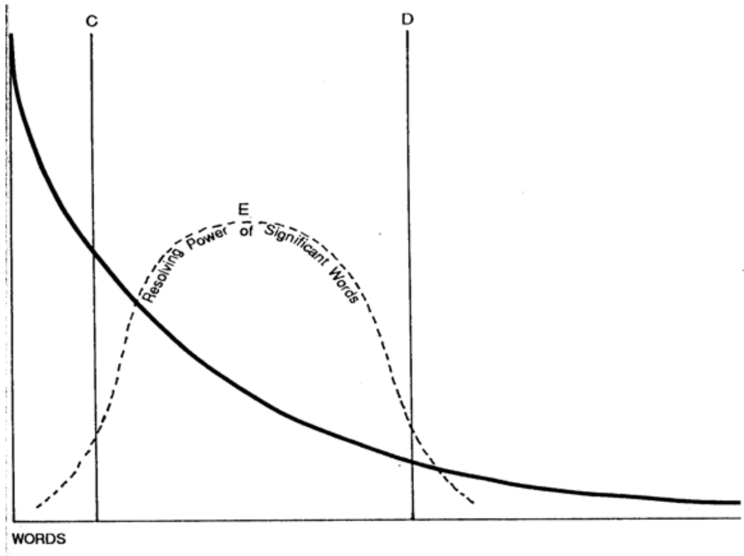

Figure 21: Luhn's theory of "resolving power"

Regarding the sample of this study, the most frequent term is "research", which occurs 833 times (see Table 7). Since "research" is a very broad term in relation to its meaning, it is understandable that many documents post to it. The explanation also applies to the numerous postings under "teaching materials", "China", "popular readings" and "basic knowledge". In fact, the term "college" belongs to this category as well, because it mostly co-occurred with "teaching materials" to indicate that the intended users of a document are college students.

Table 7: The top ten most frequent terms in the whole sample

rank	frequent terms	frequency of occurrence
1	research	833
2	teaching materials	547
3	China	473
4	college	427
5	microcomputer	289
6	popular readings	256
7	basic knowledge	246
8	plant	232
9	Flora	224
10	secondary school	213

But, why do so many items gather under “microcomputer”, “plant”, “Flora” and “secondary school” as well? Does it contradict the widespread assumption that in a database the more general vocabulary terms have more documents under them than the specific terms? In a comprehensive database, they are specific terms. However, when in a specialized database, they are the central terms of the disciplines, to which they belong. Thus, it is possible that general terms have fewer postings than specific ones. Since the sample of this research is restricted to botany, computer science, psychology and education, we can conceive that we calculate the frequencies of these terms from four small specialized databases. This is why the four specific terms have so many postings.

3.3.3.2 Rank-frequency distribution of index terms

In real and virtual worlds there are phenomena where large events are rare, but small ones quite common. The well-known instance is Zipf’s law, which states that, in a corpus of natural language utterances, the frequency of any word is roughly inversely proportional to its rank in the frequency table. That is, when computing word frequencies and ranking the words in decreasing order of their frequency, the frequency of a given word multiplied by its rank order is approximately equal to the frequency of another word multiplied by its rank (Zipf, 1949).

Actually, Zipf's law applies not only to natural language, but also to artificial language—indexing language. Figure 22 shows the rank-frequency distribution of index terms. The distribution is skewed, and the curve has the familiar Zipfian shape. One can observe that a few terms occur more than 100 times, whereas most terms occur less than 100 times (741 out of 2,220 terms occur only once). The infrequent terms cumulatively outnumber the initial portion of the graph, such that in aggregate they comprise the majority, which is demonstrated by the long tail in Figure 22.

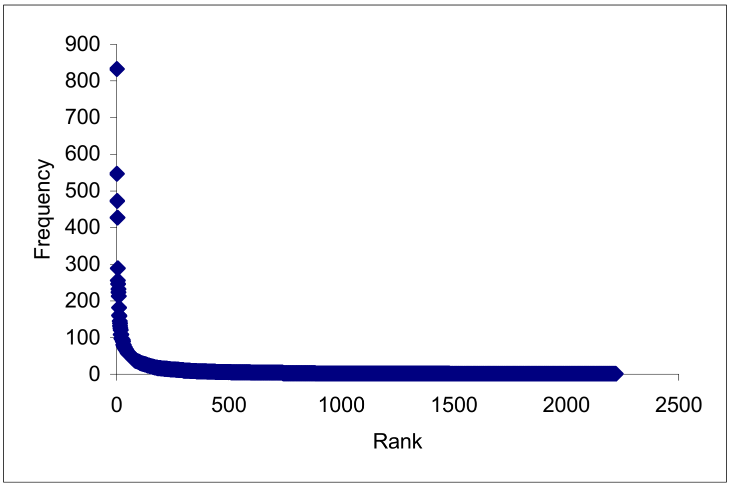

Figure 22: Linear scale plot of the rank-frequency distribution of index terms

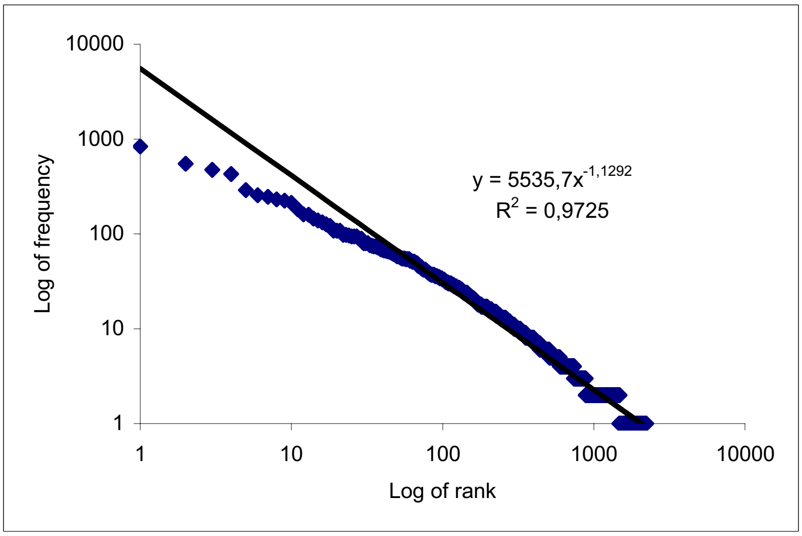

Figure 23: Log-log scale plot of the rank-frequency distribution of index terms

Figure 23 shows the same plot, but on a log-log scale. The same distribution shows itself to be nearly linear on a log-log plot, and the slope is -1.13. It indicates that the rank-frequency distribution of index terms does not strictly follow the classic Zipfian form. The initial portion of the curve deviates from this straight line. That is, the largest frequencies do not conform to Zipf's law, and they would be over-estimated using Zipf's law. In order for there to be perfectly linear relationship, the most frequent terms would have to be more frequent, and the less frequent terms slightly more numerous.

Actually, the deviation of the initial portion may be resulted by vocabulary control. When a thesaurus is constructed, the specificity of the vocabulary should be optimized. Usually, it should not contain too many non-specific terms. Meanwhile, since less specific terms are less effective as a means of retrieval, libraries usually have some indexing guidelines to restrain indexers from indexing too general terms. This is why the actual highest frequencies are much lower than the expected ones.

In order to show it more clearly that the deviation is a result of vocabulary control, we can break the whole data set into two ranges. The first 100 most frequent terms can be grouped into the first range, all the other terms into the second range. For both of the ranges we compute an idealized rank-frequency distribution. Then we get two equations for first and second ranges: $y = 1003x^{-0.73}$ and $y = 5530x^{-1.11}$, which generate two straight lines on the log-log plots, whose slopes are -0.73 and -1.11 respectively (see Figure 24). The slopes of the two lines indicate that the logarithms of frequency decreases less slowly with the logarithms of its rank in the first range than in the second range. Meanwhile, from Figure 24 we can see two forces, which are illustrated by the two light arrows. The first force (in the vertical direction) tends to reduce the frequencies of most frequent terms, whereas the second force (in the horizontal direction) tends to reduce the number of infrequent terms.

According to Zipf there are two opposing forces in the development of a language, as a result of which a certain vocabulary balance will eventually be reached. The first force is the force of unification, which tends to reduce the vocabulary and corresponds to a principle of least effort seen from the point of view of the speaker, whereas the second force, namely the force of diversification, which tends to require a large vocabulary of specialized words to describe the same concepts from the point of view of the listener (Harremoës and Topsøe, 2002). These two forces are illustrated by the two black arrows in Figure 25.

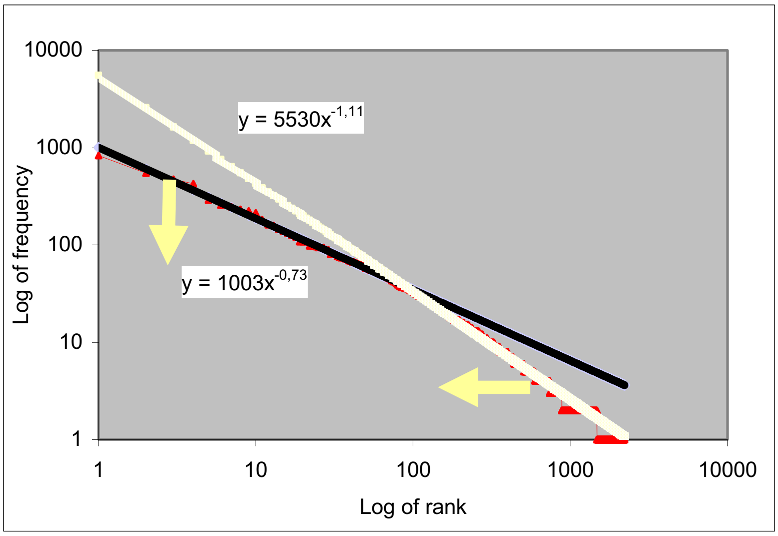

Figure 24: Log-log scale plot of the idealized rank-frequency distribution of index terms

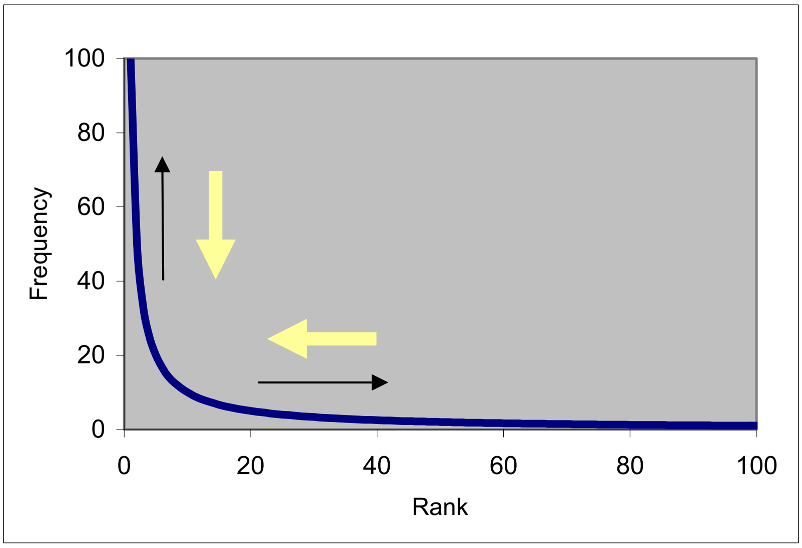

Figure 25: The forces of unification and diversification in natural language, and the control of the two forces in indexing languages

Zipf's law is originally concerned with the natural language. Considering artificial languages--indexing languages, the rank-frequency distribution of index terms does not strictly follow the classic Zipfian form. The deviations can be explained by the characteristics of indexing languages. Actually, the purpose of indexing languages is to control and balance the two forces in the natural language. On the one hand, when a controlled vocabulary is constructed, very common terms (frequently used in natural language) are left out, because they are usually useless in distinguishing document sets. Through this kind of process, the force of unification to some extent is controlled. On the other hand, synonyms and homonyms are controlled, so that every object or concept would be designated by only one term, and each term would refer to only one concept or object. At the same time, the very specific terms are excluded, because they do not serve to cluster documents in a useful way. In this sense, the force of diversification is controlled. The control of the two forces is shown by the two light arrows in Figure 25.

3.3.3.3 Recalculation of indexing consistency at the term level

In order to analyse the relationship between indexing consistency and term frequency, we should look at indexing consistency in a different way—at the term level. That is, we can measure the degree to which a term is consistently assigned in two databases. On the one hand, we can define the indexing consistency of a term as the proportion of paired records under it. For instance, in the sample from two databases, a total of 232 records are indexed with the term “plant”, and 178 of which are in pairs. The other 54 records are single. The proportion of paired records is 178/232*100=77%.

One the other hand, we can define the indexing consistency of a term as the proportion of consistently indexed books under it. For the same example we mentioned above, we can conceive it in another way. The 54 single records stand for 54 different books, while the 178 paired records for 89 different books. That is, the total number of distinct books indexed with the term “plant” is 143. Since 89 out of 143 books are indexed consistently, the consistency is 89/143*100=62%.

Here, we use the first method to recalculate indexing consistency at the term level. The overall consistency is 49% in the light of the total 2,220 distinct terms. If we leave out the 741 terms with only one occurrence, the overall consistency is 74%. In the following texts the 741 terms with only one occurrence are called “single terms”.

3.3.3.4 Relationship between consistency and term frequency

As mentioned above, over a third of the terms (741 out of 2,220 terms) are single terms. Should we leave out the single terms when we discuss the relationship between consistency and term frequency or not? On the one hand, these single terms occur in this sample because of the relative small sample size, which will not happen with regard to the whole population. Thus, the single terms are not representative or even result in bias if we include them, because the consistency of all the single terms is zero, which means that the including of the single terms would reduce the mean consistency of infrequent terms. On the other hand, it may be dangerous if we simply leave them out, considering the big proportion (a third of the whole sample). In this sense, it would be appropriate to discuss the relationship between consistency and term frequency in two different cases, i.e. leaving out and including the single terms.

If we leave out the single terms and calculate the Pearson's correlation coefficient *r*, we find that the value of *r* is 0.002, which almost equals to zero. It means that there is no linear relationship between consistency and term frequency in the first case. With reference to Figure 26 we can see that the coefficient of determination, R^2, is close to zero, which indicates that only 1.18% of the variance of *consistency* can be "accounted for" by changes in *frequency* and the linear relationship between *consistency* and *frequency*. In other words, the small value of R^2 suggests that the linear model is a poor model for the two variables (*consistency* and *frequency*).

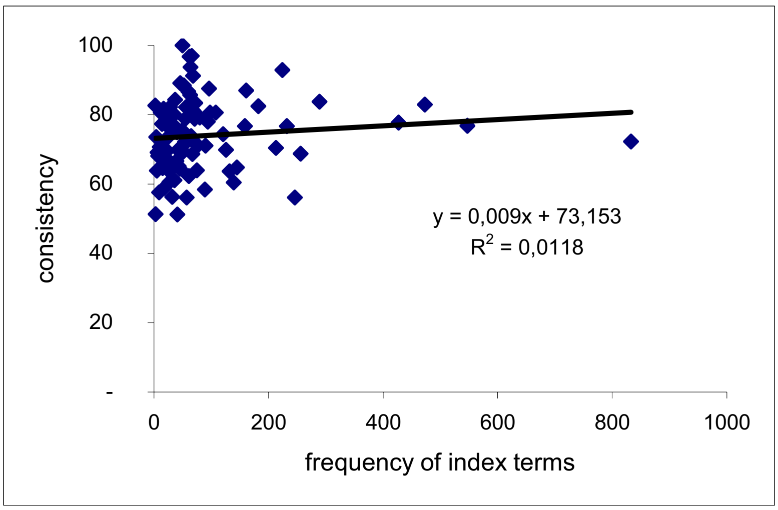

Figure 26: Relationship between consistency and frequency of index terms (leaving out single terms)

If we conceive the first hundred terms in the ranking of term frequency as frequent terms and the others as infrequent terms, we find that the consistency of infrequent terms is not as high as we thought it would be, and the consistency of very frequent terms is not what we expected as well.

If we include the single terms and calculate the Pearson's correlation coefficient *r,* we can find that the value of *r* is 0.13, which is larger than the value we get in the first case. Nevertheless, it is too small to conclude that there is a linear relationship between consistency and term frequency. The linear regression also shows that the linear model does not fit (see Figure 27).

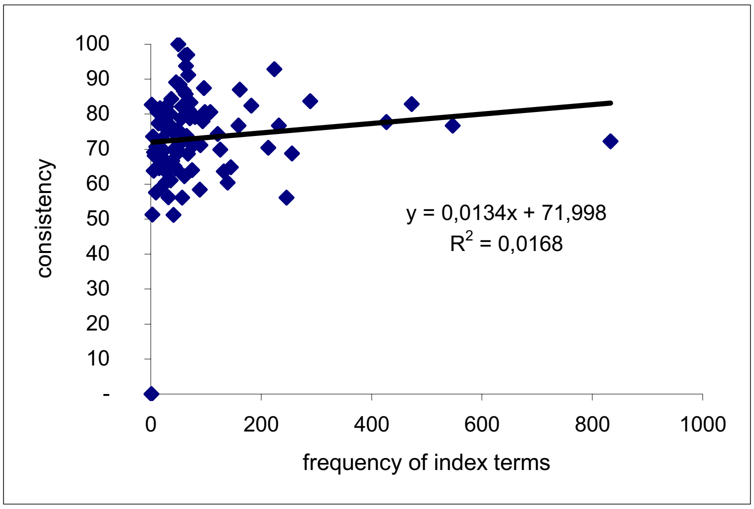

Figure 27: Relationship between consistency and frequency of index terms (including single terms)

To sum up, there is no linear relationship between consistency and frequency of index terms. In other words, greater specificity does not necessarily bring about lower consistency.

3.3.4 Length of monographs indexed

Another factor that may affect indexing consistency is the length of items indexed. In the context of this research, it is the number of pages of the monographs. An overview of the length of the monographs is shown in the following sections.

3.3.4.1 Frequency distribution of the number of pages

As described above, a total of 3,307 monographs are involved in this sample. These books differ significantly in length, which ranges from 28 pages to 2,835 pages. In the light of the big variability of length, they are grouped into 11 categories. The books with 1-99 pages are considered as the first category, the books with 100-199 pages as the second category, the books with 200-299 pages as the third category, and so on. Similarly, the books with 900-999 pages belong to the tenth category. But the books with over 1,000 pages are considered extremely long, and all of them are combined into the eleventh category.

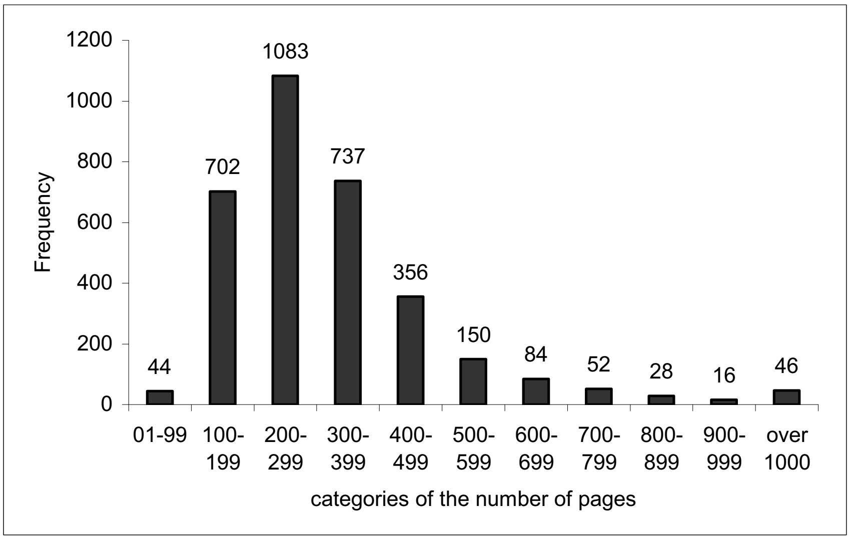

Figure 28: Frequency distribution of the number of pages

Figure 28 shows that 32.84% (1,083 out of 3,298) of the books belong to the third category (with 200-299 pages), and 76.47% of the whole sample (2,522 out of 3,298) falls into the range of 100-399 pages. Note that the total number 3,298 here is not identical to the total number 3,307 of the whole sample. The reason is that for some books the information about the number of pages is missing. Since the analysis is based on the bibliographical records rather than the books themselves, the records without pages information must be eliminated.

3.3.4.2 The relationship between average number of index terms and average number of pages

Generally, as the number of the pages increases, so does the average number of index terms assigned. "Since longer documents are capable of discussing more topics, they are capable of being about more. Longer documents are more likely to be associated with more keywords, and hence more likely to be retrieved."(Belew, 2000) However, this is not always the case. Figure 29 shows that the average number of index terms does not vary much with the increase in the number of pages. The linear regression equation y = 0.0622x + 5.266 indicates a positive correlation between the average number of index terms and the number of pages. However, the number of index terms does not change much, as the number of pages increases, which is implied by the small slope of the regression line (0.0622). T-tests show that only the two smallest values (5.39 and 5.39) and the two largest values (6.61 and 5.88) are significantly different from the overall mean value ($p < .05$).

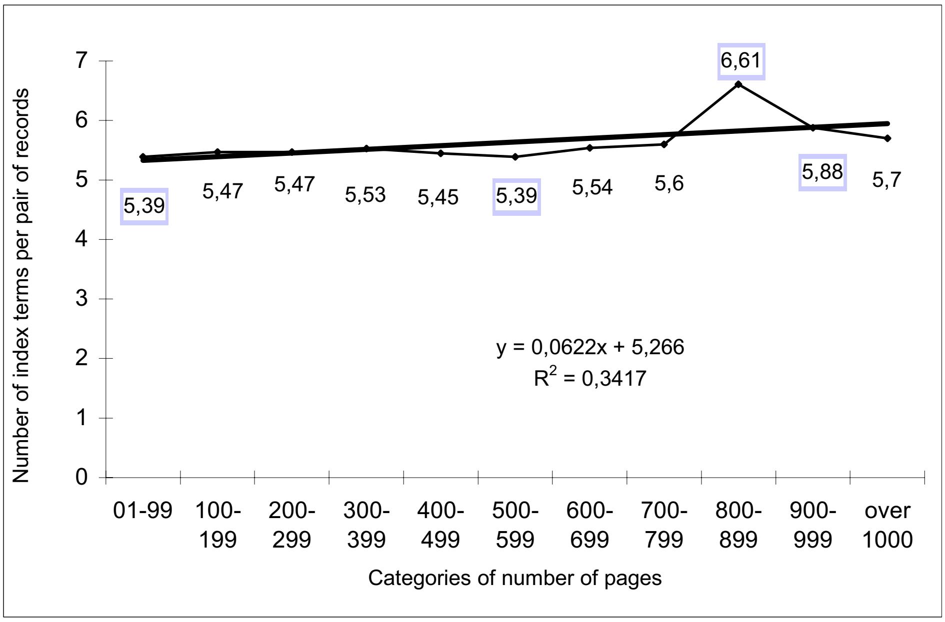

Figure 29: Relationship between the average number of index terms and the number of pages

Since the numbers of books in the eleven categories are considerably different, it may be dangerous to simply compare the mean values, especially when very few books are contained in one category, because they may be not representative enough of the category, to which they belong. In order to avoid a bias of this kind, the whole sample is divided into two parts. One part comprises the books with 100-999 pages, and the other part comprises the extremely short (<100 pages) and the extremely long (>1000 pages) books.

Firstly, if we only examine the extremely short and the extremely long books, we can find that there is almost no difference in the number of index terms between them, which is indicated by Figure 30. Moreover, a t-test is conducted, and the result also shows no statistically significant difference between the extremely short books and the extremely long books concerning the number of index terms per book.

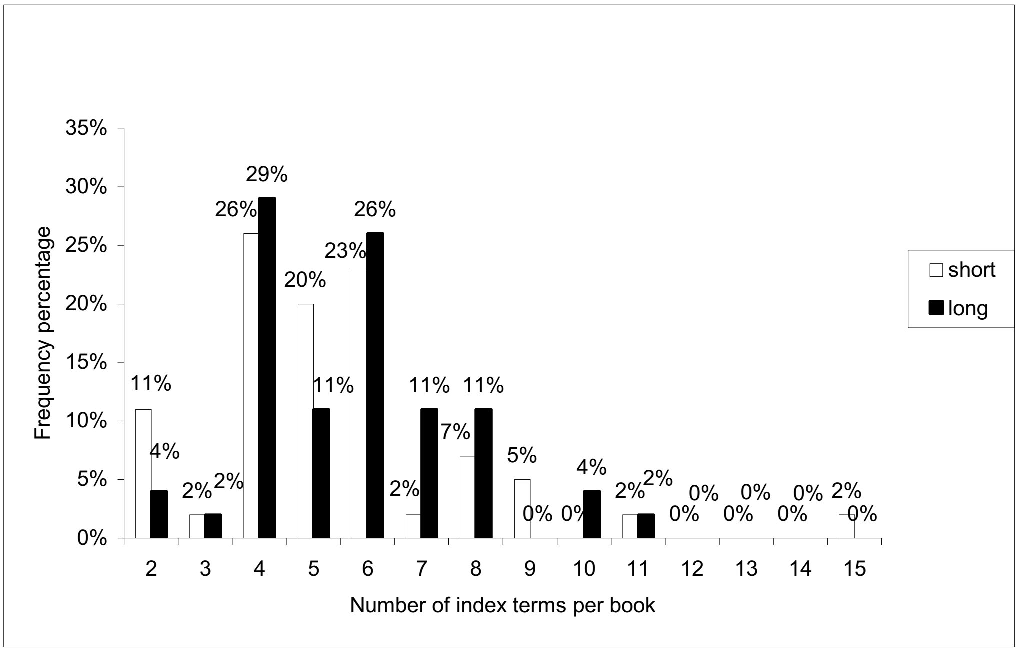

Figure 30: Comparison of the extremely short and extremely long books concerning the number of index terms per book

Then, we can look at the other part of the whole sample, i.e. the books with 100-999 pages. Since too few books fall into the seventh (namely 600-699), eighth (700-799), ninth (800-899) and tenth (900-999) groups, which contains 84,52,28,16 books, respectively (see Figure 28), we should unite these four groups into one group. Based on the re-categorizing, we can notice that without extremely short and the extremely long books the relationship between the number of index terms and the number of pages does not change much. Table 8 indicates that the values of the number of index terms are almost the same in relation to different categories, although there is a very slight increase.

Table 8: Relationship between the number of index terms and the number of pages

Categories	Range of pages	No. of books	Mean number of index terms	Standard deviation
1	100-199	702	5.47	2.22
2	200-299	1083	5.47	2.23
3	300-399	737	5.53	2.14
4	400-499	356	5.45	2.02
5	500-599	150	5.39	2.39
6	600-990	180	5.57	2.14

Obviously, Lancaster's statement: "the shorter the item, the fewer the terms that might plausibly apply" doesn't hold for the data of this research (Lancaster, 1998). In fact, based on this sample, we can conclude that length of books has no profound influence on indexing exhaustivity. It seems that indexing exhaustivity depends more on indexing policies or the contents of documents in hand than on length of them.

3.3.4.3 The relationship between length of monographs and indexing consistency

It is commonly believed that indexing consistency increases as documents decrease in length. However, for the whole sample, a Pearson correlation shows no statistically significant correlation between length and consistency.

Again, we divide the whole sample into two parts. One part comprises the books with 100-999 pages, and the other part comprises the extremely short (<100 pages) and the extremely long (>1000 pages) books. The books of the first part are grouped into six categories (see Table 9)

Table 9: Categories of the number of pages ranging from 100-999

Category	Range of pages	No. of books
1	100-199	702
2	200-299	1083

3	300-399	737
4	400-499	356
5	500-599	150
6	600-999	180

Excluding the extremely short and extremely long books a Pearson correlation shows the same result, i.e. no statistically significant correlation between length and consistency. Figure 31 shows the indexing consistency of different categories. And there is no evidence showing that as the number of pages increases, so does the indexing consistency.

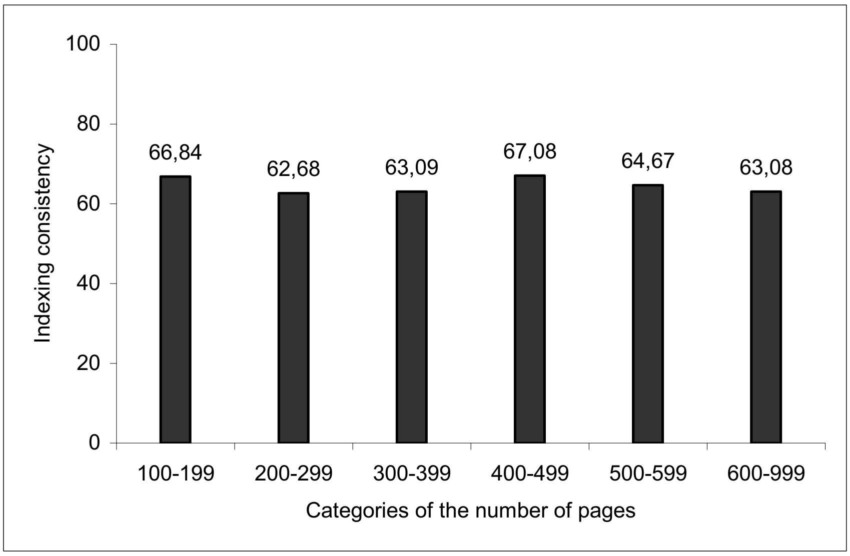

Figure 31: Chart of indexing consistency in relation to different categories of the number of pages

For the second part of the sample, a t-test is conducted, and the result also shows that although the mean consistency of extremely short books (63.3%) is a little higher than that of the extremely long books (60.73%), the difference is not statistically significant.

To conclude, there is no obvious evidence showing a correlation between length and consistency.

3.3.5 Category of indexing languages

Another thing to note is that the type of vocabularies can affect indexing consistency as well. However, researchers are not consistent with each other in categorizing vocabularies. Slamecka and Jacoby (1963) divided vocabularies into two categories, namely "prescriptive" and "suggestive". And they claimed that using prescriptive vocabularies could significantly improve indexing consistency, because prescriptive vocabularies gave indexers no semantic freedom by term choosing. Meanwhile, they stated: "Quality of indexing is best improved by vocabularies which formalize relationships so as to uniformly and invariably prescribe the choice of indexing terms." It sounds reasonable.

Vocabularies can be divided in a totally different way, for instance, they can be categorized into subject headings, thesauri and classification schemes. Markey (1984) summarized several consistency studies and concluded: "A pattern emerges showing that greater interindexer consistency is attained when indexers employ classification schemes." It sounds plausible and questionable. Why should indexing consistency be higher when classification schemes are employed? Does it hold for the data of this research? In order to answer these questions, the author compares indexing consistency of index terms and that of class numbers and finds that the mean consistency of index terms is obviously higher than that of class numbers, which is shown in Table 10. And a t-test shows that the difference is significant ($p<.01$). Obviously, Markey's conclusion does not hold for the data of this study. That is to say, indexing consistency is not necessarily higher when classification schemes are employed.

Table 10: Comparison of indexing consistency between index terms and class numbers

Statistics	consistency of class numbers	consistency of index terms
Mean	61.58	64.21
Median	100.00	66.67
Mode	100.00	100.00
Std. Error of Mean	0.80	0.64
Std. Deviation	45.82	37.04
Skewness	-.47	-.41
Kurtosis	-1.65	-1.35

Actually, the result testifies Slamecka and Jacoby's theory, because the Chinese Library Classification is relatively more suggestive than the Chinese Thesaurus. The Chinese Library Classification falls between enumerative and faceted classification. Many subjects need to be synthesized by the number-building process, which easily results in indexing inconsistency. Moreover, there are many alternative classes in the classification scheme, which are used not by force of the classification system, but are decided by the library's indexing policy or by the indexer. The flexibility of using the alternative classes to some extent causes the inconsistency. In comparison with the Chinese Library Classification, the Chinese Thesaurus does not give indexers much semantic freedom by term choosing.

In previous consistency studies, only the difference between thesauri and classification schemes has been discussed. No one has investigated whether there is a correlation between consistency of index terms and consistency of class numbers. After calculating Pearson's correlation coefficient (r=0.42), we can see that there is a positive linear correlation between consistency of index terms and consistency of class numbers. However, if we examine the relationship more closely, we can find that it is not a strict linear correlation.

When the consistency of class numbers is 0%, the average consistency of index terms is 44.05% rather than 0%. When the consistency of class numbers is 100%, the average consistency of index terms is 77.33% rather than 100%. When the consistency of class numbers is larger than 0% and smaller than 100%, there is a logarithmic increase in consistency of index terms, as the consistency of class numbers increases (see Figure 32). Similarly, when the consistency of index terms is 0%, the average consistency of class numbers is 25.82%. When the consistency of index terms is 100%, the average consistency of class numbers is 81.05%. When the consistency of index terms is larger than 0% and smaller than 100%, the consistency of class numbers increases logarithmically, as the consistency of index terms increases, although the logarithmic curve doesn't fit as well as the curve in Figure 32 does (see Figure 32 and Figure 33).

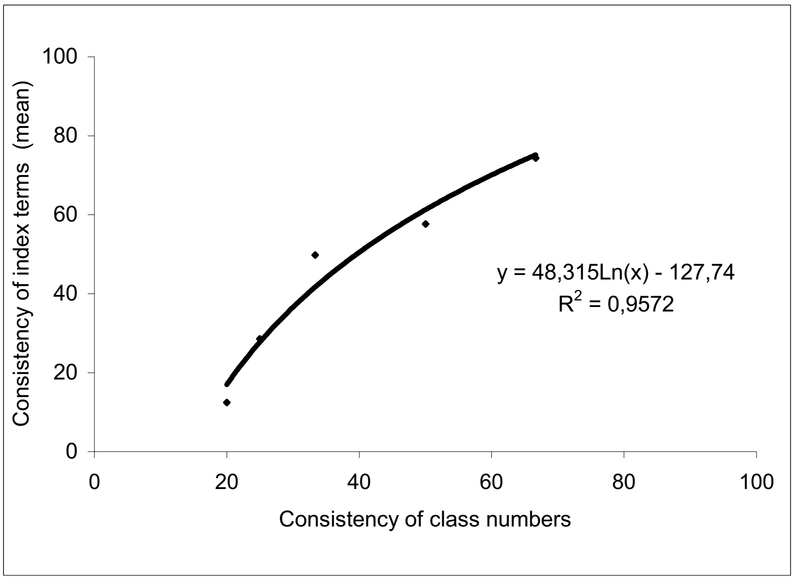

Figure 32: Relationship between consistency of class numbers and consistency of index terms (where consistency of class numbers is larger than 0% and smaller than 100%)

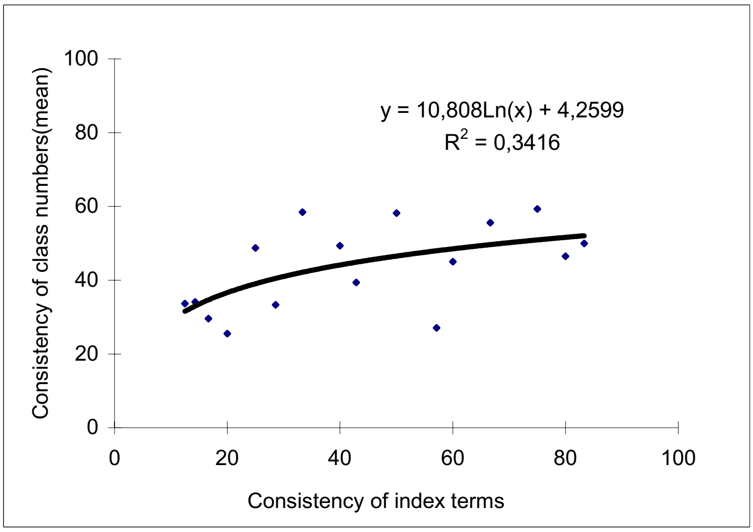

Figure 33: Relationship between consistency of index terms and consistency of class numbers (where consistency of index terms is larger than 0% and smaller than 100%)

It is not surprising that a positive correlation between consistency of index terms and that of class numbers is obtained, because the Chinese Classified Thesaurus is utilised as an indexing aid tool by both library systems. It is an integration of the Chinese Library Classification and the Chinese Thesaurus, that is to say, a two-way corresponding list between class numbers and descriptors. When indexers choose one class number, they can get suggestions of correspondent descriptors at the same time, or vice visa. Thus, when two indexers choose the same class number, they are very likely to choose the same descriptors, which are suggested in the corresponding list. Similarly, when two indexers choose the same descriptor, they are very likely to choose the same class number. However, since the Chinese Classified Thesaurus is just an aid tool, indexers can freely consult it when they need to or when they like to. Therefore, sometimes, when two indexers choose the same class number, they may choose different descriptors. This is why the consistency of index terms is not always the same as the consistency of class number, or vice visa.

Therefore, we can conclude that indexing consistency is not necessarily higher when classification schemes are employed and that using suggestive vocabularies can be detrimental to indexing consistency. Moreover, there is a positive correlation between the consistency of index terms and the consistency of class numbers. That is to say, when the consistency of index terms is high, the consistency of class numbers is very likely to be high, vice versa.

3.3.6 Subject area

A subject area's terminology can be one of the factors affecting indexing consistency. In former studies, some researchers have noticed the difference between "soft science" and "hard science" in relation to consistency. Tibbo (1994) claims that the terminology in the humanities tends to be imprecise and authors may prefer to make their writing "dense" rather than readable. Lancaster (1998) assumes that "greater consistency will occur in the indexing of more concrete topics (e.g., physical objects, named individuals) and that consistency will decline as one deals increasingly with abstractions."

The consistency among the four subject areas is compared. Table 11 summarizes the consistency values of the four subject areas, which are chosen intentionally for comparison. The result confirms Lancaster's assumption. It can be seen from Table 11 that the greatest indexing consistency occurs in the indexing of books in botany. It can possibly be explained by the fact that the terminology in this area is relatively stable, few new terms are added each year, and authors possibly use the older terms more consistently.

Table 11: Comparison of indexing consistency among the four subject areas

domain	No. of books	consistency of index terms	consistency of class numbers
computer	850	56.43	54.59
education	850	62.84	60.99
psychology	848	67.37	65.31
botany	759	70.87	65.89

ANOVA shows that consistency is statistically significant different among the four subject areas.

Further more, we can see from Table 11 that indexers are less consistent with each other in indexing the books in psychology and education than in botany. In comparison with botany, the terminology of psychology and education can be conceived of as “soft”, because psychology and education belong to humanities and social science, respectively. Thus, it is understandable that lower consistency occurs in these two areas.

It is surprising that the lowest consistency occurs in computer science. Computer science cannot be regarded as “soft science”. It belongs to the area of technology. Why does the lowest consistency occur in this subject area comparing with the other three domains? The most important reason should be that this discipline develops so quickly that the revision of indexing languages cannot keep pace with its development. And the terminology of this area is unstable. Various new concepts come out very rapidly. Authors are inclined to use new terms to show that they have some new ideas or find new technology. The delayed revision of indexing languages makes both indexing and retrieving difficult. Just as Samleske (1992) complained, “On the whole, information on computer software can be accessed effectively in terms of currency and completeness, but the procedure is time-consuming due to a large number of competing headings and an unclear hierarchical structure”.

3.3.7 Development of disciplines

In Subsection 3.3.6, we have discussed how the terminology of a discipline affects indexing consistency. However, the terminology of a discipline is closely related to the development of this discipline. As mentioned above, since botany develops very slowly, its terminology is relatively stable. In contrast, the quick development of computer science causes the unstable terminology, which contributes to the low consistency. In this sense, the development of a discipline can indirectly influence indexing consistency.

In order to obtain an insight into the development of the four subject areas, the number of books in National Library of China in different years is observed. Figure 34 shows that the number of books in computer science increases most dramatically among the four domains. However, the number of books in botany varies least over the years. It is notable that the more quickly a discipline develops, the more difficult it is to achieve high consistency. Moreover, we can see from Figure 34 that before 1996 the number of books decreases, after 1996 it increases. We discuss them separately in the following sections.

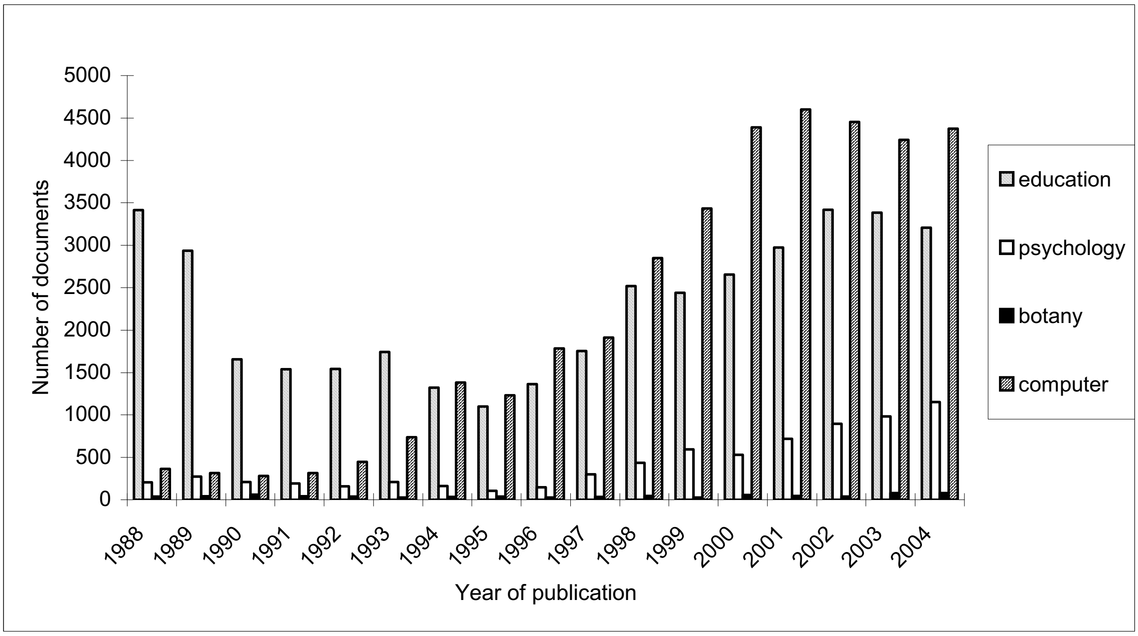

Figure 34: The development of documents in the four domains

3.3.7.1 The change of the number of books in the four subject areas from 1988 to 1995

By referring to Figure 35 we can see that the total number of books in the four subject areas declines from 1988 to 1995. It is mainly because of the inflated price of books and libraries' shortage of money. On the one hand, more and more books were published and put on the market during this period. On the other hand, books became more and more expensive. For libraries, it was very difficult to collect all the books on the market. Due to the shortage of money, many libraries even cut their budget for new books, which caused the decrease in the total number of books that they collected over the years. The conflict between the increase of books' price and the decrease of libraries' budget was the biggest problem that libraries had to face. And at that time this issue was a focused topic in the literature of this field.

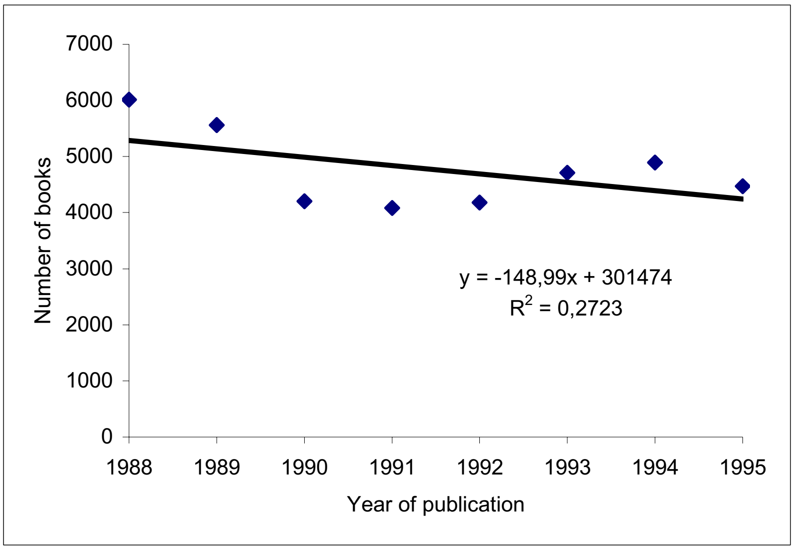

Figure 35: The decrease in the number of books indexed per year from 1988 through 1995

Upon comparison of Figure 36, Figure 37, Figure 38 and Figure 39 we find that although the total number of books declines, the number of books in computer science is on the rise. Since computer science was a new subject area at that time, the demand on the books in this subject area was great. Thus, many books in computer science were bought, although the library had no much money.

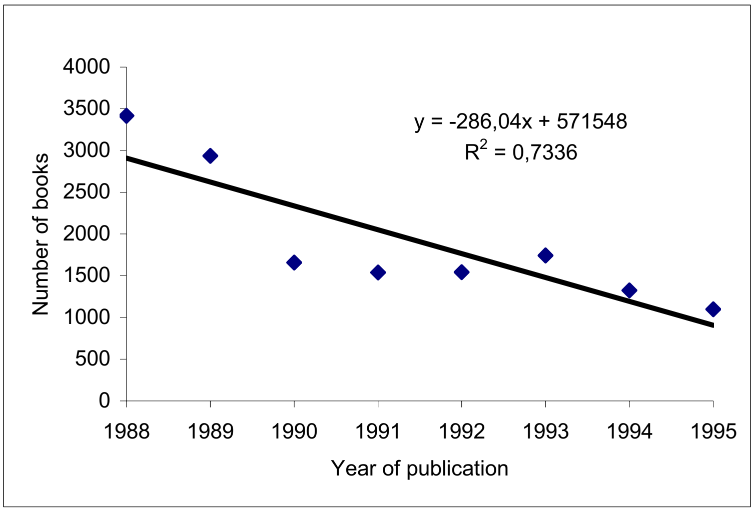

Figure 36: The decrease in number of books indexed per year from 1988 through 1995 (education)

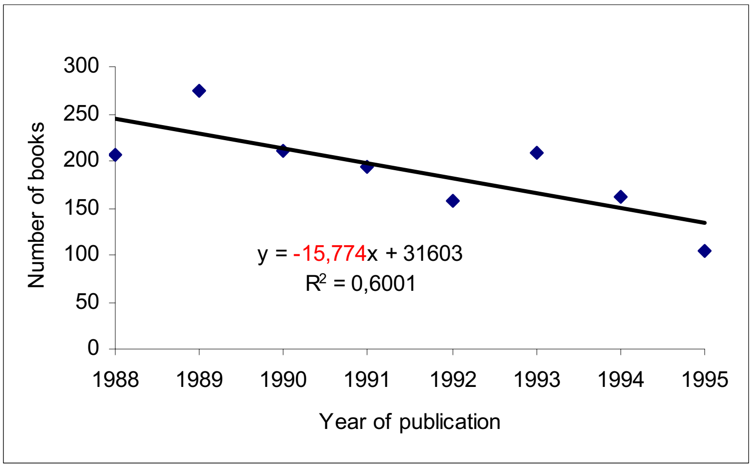

Figure 37: The decrease in number of books indexed per year from 1988 through 1995 (psychology)

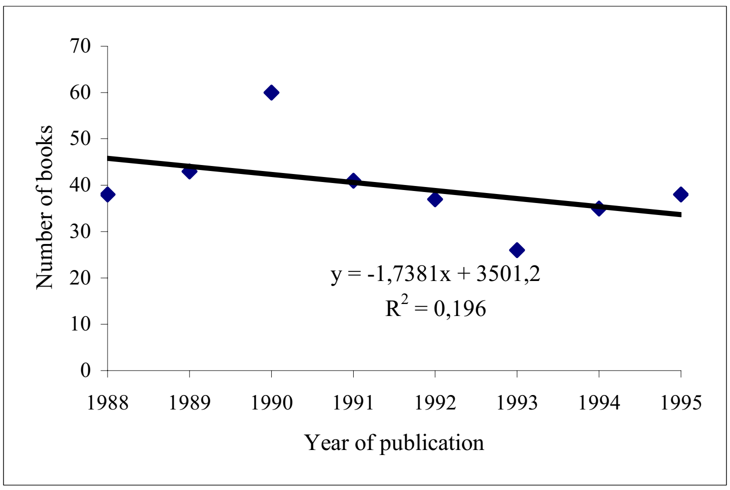

Figure 38: The decrease in number of books indexed per year from 1988 through 1995 (botany)

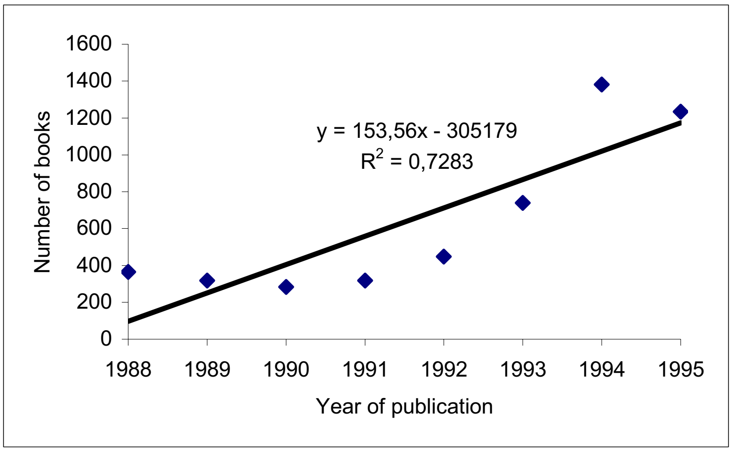

Figure 39: The increase in number of books indexed per year from 1988 through 1995 (computer science)

If we thoroughly examine the four regression lines in Figure 36, Figure 37, Figure 38 and Figure 39, we can notice that the slopes of the four regression lines vary considerably. In fact, the bigger the absolute value is, the more quickly the number of books changes. This is the meaning of the slope of a regression line.

We can see it more clearly, if we tabulate the values of consistency and the slopes, and then make a comparison of them. The second column in Table 3.11 is the average value of indexing consistency from 1988 to 1995. The third column is the slope of the regression line. The bigger absolute value of the slope is, the more quickly the number of books changes. It is evident from Table 3.11 that there is an inverse relationship between the value of consistency and absolute value of slope. It means that the more quickly a discipline changes, the more difficult it is to achieve high consistency when indexing the books in this discipline.

Table 12: Relationship between development of discipline and consistency from 1988 to 1995

Domain	Consistency (mean value)	Slope of a regression line (absolute value)
Education	65.43	286.04
Computer science	65.67	153.56
Psychology	72.58	15.77
Botany	73.18	1.74

3.3.7.2 The change of number of books in the four subject areas from 1996 to 2004

Inspection of Figure 40 indicates that the total number of books increases very quickly from 1996 through 2004. It is not because of the decline in books' price, but due to the union catalogue. In order to resolve the conflict of lack of money and the increase in the price of books, many libraries began to cooperate with each other by union cataloging in 1997. Both of the catalogs in question were involved in union cataloging. Actually, CALIS itself was a result of union cataloging. It was started in 1998. And the Online Library Cataloging Center (OLCC for short) was established in 1997, the core of which was NLC. Nowadays OLCC and CALIS are the two main systems for online cataloging in China (Liu and Shen, 2002). By union cataloging, libraries could share resources and reduce costs. Therefore, the number of bibliographical records increases in this time period.

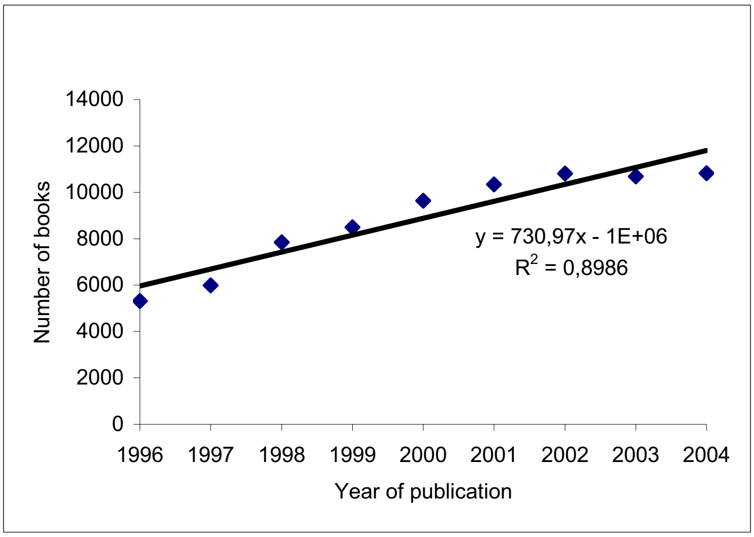

Figure 40: The increase in number of books indexed per year from 1996 to 2004

The increase in the number of books in the four subject areas is shown in Figure 41, Figure 42, Figure 43 and Figure 44. Comparison of these Figures shows that the absolute values of the four regression lines' slopes are also very different. The slope of regression line of computer science is 362, while of botany it is only 5.58. It indicates that computer science develops most quickly and that botany is most stable.

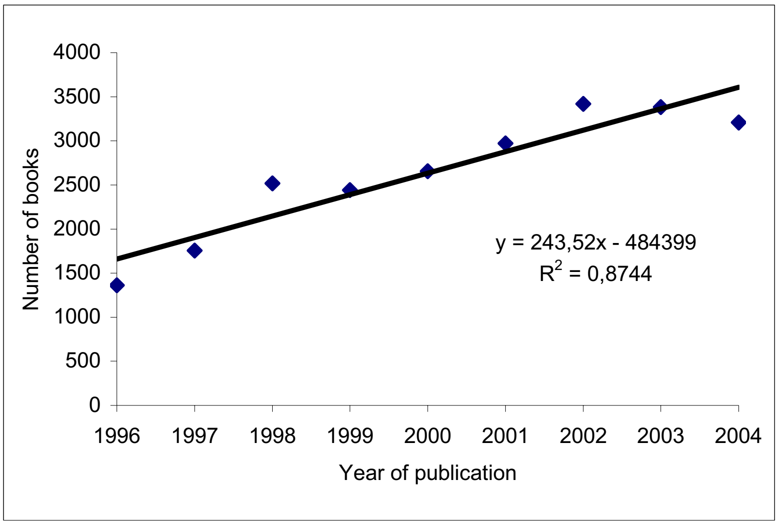

Figure 41: The increase in number of books indexed per year from 1996 through 2004 (education)

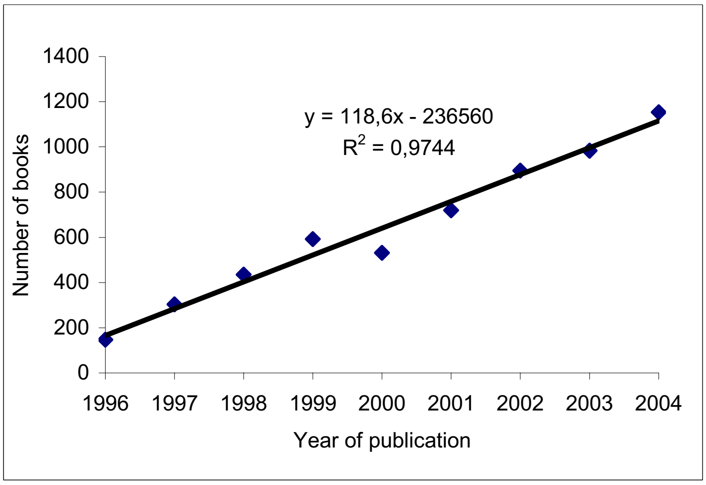

Figure 42: The increase in number of books indexed per year from 1996 through 2004 (psychology)

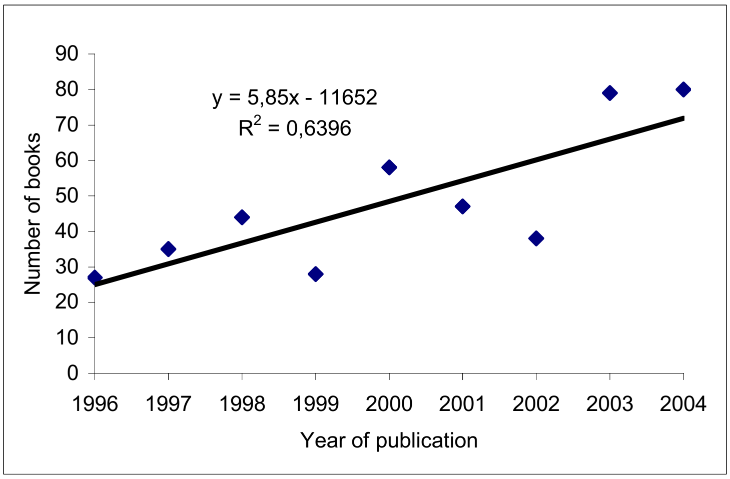

Figure 43: The increase in number of books indexed per year from 1996 through 2004 (botany)

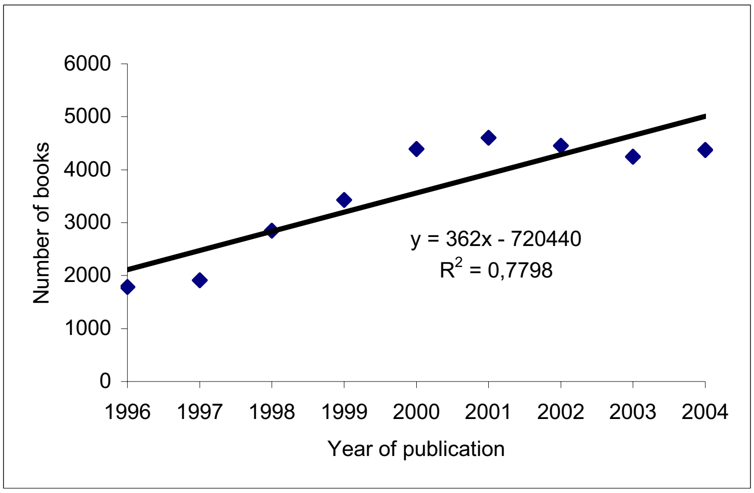

Figure 44: The increase in number of books indexed per year from 1996 through 2004 (computer science)

Again we can tabulate the value of the consistency and the slopes, and then make a comparison of them. Similar to Table 12, the second column in Table 13 is the average value of indexing consistency from 1996 to 2004. The third column is the slope of the regression line. It is evident from Table 13 that there is an inverse relationship between consistency and slope. It means that the more quickly a discipline develops, the more difficult it is to achieve high consistency.

Table 13: Relationship between development of discipline and consistency from 1996 to 2004

Domain	Consistency (mean value)	Slope of a trend line (absolute value)
Computer science	48.91	362
Education	60.55	243.52
Psychology	62.69	118.6
Botany	68.87	5.85

To summarize, the dramatic change or development of a discipline may lead to the unstable terminology, and thereby may result in low consistency in the indexing of the books in this subject area.

3.3.8 The structure of vocabulary

In Subsection 3.3.7 we have discussed that the rapid development of a discipline can lead to low consistency. In fact, to some extent it can be improved if the revisions of vocabularies can keep pace with the development of disciplines. However, it can be even worse if vocabularies cannot be revised timely, because new concepts and terminologies can not be included. When indexers need to describe a book with new concepts or topics and cannot find appropriate index terms or classes, they can only choose some relevant but not precise terms or classes, which may easily lead to inconsistency. Thus, whether the vocabularies used are up-to-date can influence indexing consistency as well. In the following text, how the unreasonable structure of vocabularies causes low consistency is analyzed by using the subject area "computer science" as an example.

As discussed above, the terminology of computer science is not as abstract as that of education and psychology. Nevertheless, the lowest consistency occurs in the indexing of books in computer science. One of the reasons is that computer science develops most quickly in comparison with the other three subject areas. The other reason for the lowest consistency is the time lag of classification scheme and controlled vocabulary in this subject area with regard to their revisions. With the development of disciplines, classification schemes and controlled vocabularies should also be revised. However, the classification scheme and thesaurus involved in this study are revised almost every 10 years. Obviously, it cannot fit the quick development of disciplines, especially when it is a discipline as computer science.

Table 14: The imbalanced distribution of postings under different classes in computer science

Number of documents under one class	Number of classes
0	48
1-100	108
101-200	4
201-300	7
301-400	6
401-500	2
501-600	3

601-700	4
701-800	1
801-900	1
901-1000	0
1001-2000	7
2001-3000	2
3001-4000	1
4001-5000	1

In reference to the classification scheme in the field of computer science, there are many out-of-date classes. From Table 14 we can see that there are 48 classes without postings. That is to say, these 48 classes have not been used for 17 years (from 1988 through 2004). There may be many reasons for it. But one important reason is that most of them are out of date. Another reason is that some classes are too specific. Moreover, it is indicated by Table 14 that a severe imbalance in postings exists among various classes. Thousands of documents gather under only several classes, whereas there are 48 classes without postings and 16 classes with only one posting. It is very interesting to notice that 64.69% of documents in computer science gather under only 5.64% of the classes. The imbalanced distribution of documents can profoundly influence classification retrieval and browse. When users retrieve with a class number or browse under a class, they can easily get zero hit or more than one thousand of records. Usually, users are frustrated by the failed retrieval and then lose their confidence in classification retrieval and browse. Further more, when a class has more than one thousand of postings, it is not discriminating any more, and thereby is not effective as a means of retrieval.

With regard to the thesaurus in the field of computer science, many terms are out of date as well. Meanwhile, many new terms are not contained in the thesaurus because of the delayed revision. Delayed revision of the thesaurus is detrimental to indexing consistency. As mentioned above, when indexers describe documents with new concepts or topics and cannot find appropriate terms in the thesaurus, they may use free terms, i.e. terms not included in the thesaurus. Earlier studies indicate that free indexing can result in low consistency. In fact, statistical result of this study shows that 13% of the indexing inconsistencies in computer science are a result of using free terms. We may say that delayed revision of the thesaurus in this field contributes to the low consistency.

3.3.9 Year of publication

The consistency values in different years are observed, and it is shown that the consistency varies over the years. On the whole, there is a downward trend over the years (see Figure 45). Moreover, from the slope (-1.0732) of the regression line we can see that the consistency declines slowly rather than dramatically over the years. One of the reasons for the decrease in the consistency is that topics in books become more and more complicated. Hence, it is more and more difficult to get consistency in indexing. Another reason may be that on the one hand many new concepts come up every year, on the other hand the thesaurus cannot supply sufficient new terms for indexing because of the delayed revision.

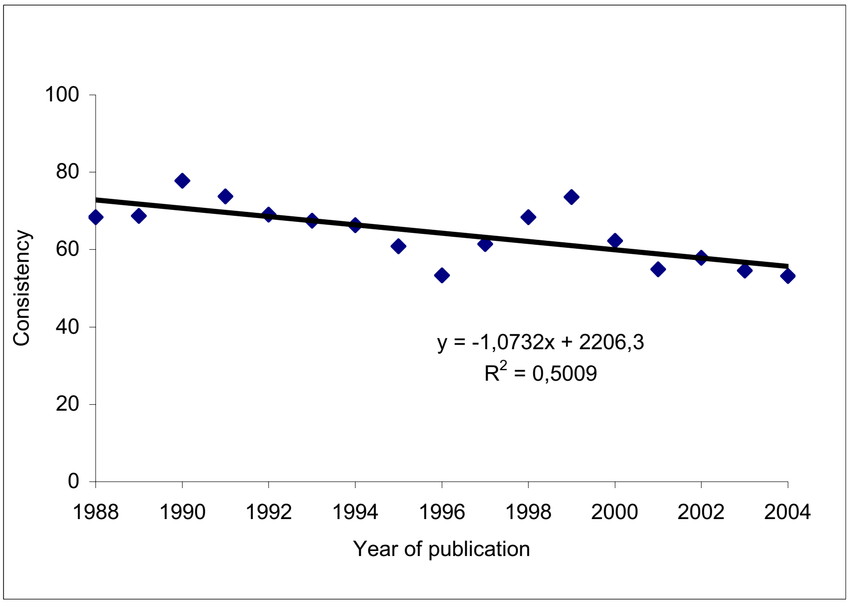

Figure 45: The relationship between year of publication and consistency

However, if we observe the graph thoroughly, we can notice that the curve is not smooth. As can be seen from Figure 45, the consistency values fluctuate from 1988 to 2004. At the beginning, the consistency holds steady and then peaks around 1990, after which there is a gradual decline. By 1996, it falls to its lowest point. Subsequently, it rises considerably to another peak in 1999, and then dips again. In the literature some evidence is found that union cataloging began from 1997 in China. Union cataloging has changed the way of indexing, because during cataloging indexers can easily search in their own or other online catalogues and directly accept other indexers' subject description, which may lead to high consistency. This is the reason for the increase in the consistency since 1997. But, on the other hand, the number of books to index is much larger than before, whereas the number of indexers is relative small. Indexers have to do more work than before, thereby they have more stress. It can make them loose their patience, which can lead to bad indexing quality. This is an explanation for the decline in consistency from 1999 to 2004.

3.4 Conceptual analysis of indexing inconsistencies

In Section 3.3 some factors influencing indexing consistency have been introduced and analyzed. However, the relationships between consistency and these factors are somewhat indirect. We still need to speculate the direct reasons for inconsistencies. In order to determine the reasons for inconsistencies, researchers usually analyze indexing mistakes, and then make some inferences based on the analysis. However, we can only say that there is bad or good rather than right or wrong indexing, we need to analyze the reasons for inconsistencies based on the analysis of the categories of inconsistencies. Further more, when we analyze the reasons for indexing inconsistencies, we should take into consideration indexers, indexing policies, controlled vocabularies and documents, because they are involved in the whole indexing process and play significant roles. In the following texts, firstly the inconsistencies are categorized, and then the distribution of these categories is examined. Subsequently, the distributions of these categories among the four subject areas are compared. Lastly, the reasons for different categories of the inconsistencies are analyzed.

3.4.1 Inconsistencies of index terms

3.4.1.1 Categories of inconsistencies

The inconsistencies are categorized based on the grouping of index terms. The terms in the Chinese Thesaurus can be divided into five categories: topical terms, form terms, geographic terms, general terms, and other terms. Topical terms deal with abstract and concrete concepts and ideas, objects, activities, processes and so on. They indicate what a work is about. For instance, “hardware” and “high school” are topical terms. Form terms can represent the bibliographic, literary, or artistic form in which the material on a subject is presented. They indicate what a work is rather than what it is about. For instance, “collections”, “dictionaries”, “encyclopedias”, “teaching materials”, “yearbooks” and so on are all form terms. General terms are some general conceptions. They don’t belong to any specific subject areas, for instance, “research”, “introduction”, “relation”, “analysis”, etc. Geographic terms are place names. It can be a country, a political division, a location and so on. For instance, “China”, “Beijing”, “Yellow River”, etc. belong to this category. Other kinds of terms include personal names, chronological terms, language names, etc.

According to the grouping of terms, the inconsistencies can be categorized as follows:
Category 1: inconsistencies in topical terms
Category 2: inconsistencies in form term
Category 3: inconsistencies in general terms
Category 4: inconsistencies in geographic terms
Category 5: inconsistencies in other terms

3.4.1.2 Distribution of different categories of inconsistencies

Among the total of 3,307 pairs of records 1,812 pairs are not exactly consistent in index terms. Figure 46 is a bar graph, depicting relative frequencies of different categories of inconsistencies. It indicates that more than 50 percent of the inconsistencies fall into the first category, i.e. inconsistencies in topical terms. It is not surprising that indexers have more difficulties in assigning topical terms than the other categories of terms, because the subject matter of a document in itself is difficult to be determined. It is a subjective process to determine what a document is about. Sometimes it is inevitable that two indexers interpret the same documents in different ways. Therefore, it is not easy to make an improvement for this kind of inconsistencies.

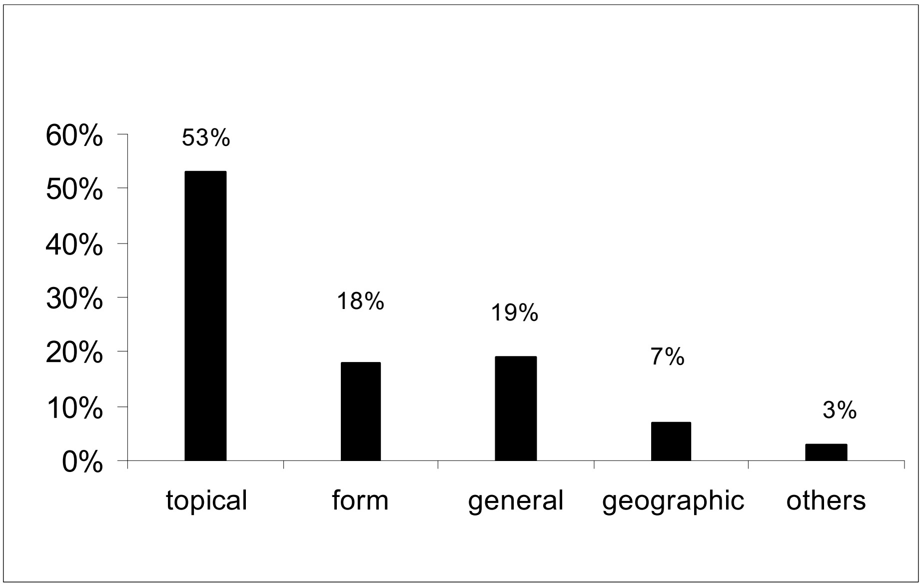

Figure 46: Relative frequency distribution of different categories of inconsistencies (index terms)

Another point shown by Figure 46 is that 18% and 19% of the inconsistencies are the inconsistencies in form terms and general terms, respectively. Generally, form terms and general terms can only represent minor subjects or minor aspects of a subject. Whether or not assigning a form term or general term becomes an arbitrary decision. Moreover, form terms and general terms are sometimes not well-defined in the thesaurus, e.g. their meanings are vague, and their intensions are unclear. This can make indexers confused, and consequently cause inconsistencies.

In comparison with topical terms, form terms and general terms, geographic terms are easier to be assigned consistently, because their meanings are usually very clear. However, there are also some problems on assigning geographic terms, especially for the subject area botany. In the following texts we will compare the five categories of inconsistencies among the four subject areas.

3.4.1.3 Comparison of the four subject areas

With respect to the five categories of inconsistencies, the relative frequency distributions are considerably different among the four subject areas, which are illustrated in Figure 47. If we closely examine every category of inconsistencies in relation to the four subject areas, we can notice that every subject area has its own characteristics and its own problems in indexing.

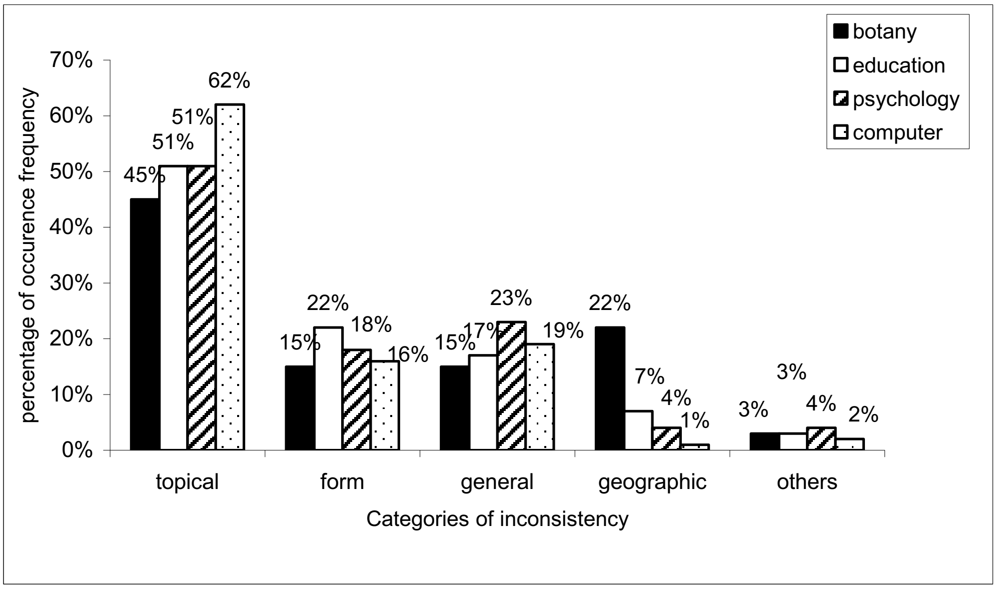

Figure 47: Comparison of different categories of inconsistencies in index terms among the four subject areas

Considering the category of inconsistencies in topical terms, it is obvious that indexers have more problems in indexing books in computer science. The proportion (62%) is larger than the overall mean value (53%). The fact that indexers have bigger problem on assigning topical terms in computer science may be due to the quick development of computer science and the delayed revisions of the vocabulary. When indexers cannot find an appropriate term to describe a new concept, they index it uncontrolled, i.e. using their own terms freely, which brings about the problem on assigning topical terms in computer science.

Referring to the category of inconsistencies in form terms, the subject area education stands out (the proportion is 22%). For the subject area education, it is important to distinguish different forms of documents, for instance, dictionaries, handbooks, teaching materials, textbooks, learning materials for exams, reading materials after school and so on. As discussed above, since they represent what a document is rather than what it is about, form terms are usually considered as minor terms. Whether or not using them is an arbitrary decision.

With regard to the category of inconsistencies in general terms, it can be seen from Figure 47 that the column of psychology is the highest (the proportion is 23%). Books in psychology are more likely to deal with some general concepts, for example, research, introduction, methods, etc. These concepts are too broad, and the meanings are not clear. Sometimes there is even some intersection between two general terms. Usually general terms are not clearly explained and isolated in the thesaurus, i.e. they have no contexts and can be used freely. Further more, whether or not using them is not by force due to their incapability of describing major subjects or major aspects of a subject.

If we look at geographic terms, we can find that there is a big problem on assigning geographic terms in the subject area botany (the proportion is 22%), while there is almost no problem in computer science (the proportion is 1%). It may be explained by the fact that many books in botany deal with geographic aspects, whereas the books in computer science do not. One of the reasons for the inconsistencies in geographic terms is lack of authority files. Different names are assigned to represent the same places. For example, full names and shortened names are used by different indexers at the same time. Another reason is that indexers are not sure when they should assign geographic terms, since geographic aspects are often not major subjects.

From the comparison we can notice that every subject area has its own characteristics and its own problems in indexing. We need more detailed analysis on every subject area, and then we can make different indexing rules for different subject areas, according to their own characteristics. By doing this, we can to some extent improve indexing consistency.

3.4.1.4 Reasons for different categories of inconsistencies

As Lancaster stated, subject indexing involves two principal steps: conceptual analysis and translation. But, the boundary between these two steps is not always clear. In fact, they may occur simultaneously. Therefore, we cannot separate these two steps and discuss them respectively. However, we can discuss the reasons based on thoroughly categorizing the inconsistencies. In the following texts, the five categories of inconsistencies mentioned above are further divided.

3.4.1.4.1 The inconsistencies in topical terms

The inconsistencies in topical terms can be further divided into the following five subcategories:

Subcategory 1: inconsistencies in terms which are synonyms or closely related, for instance, “human relationships” vs. “interpersonal relationships”, “computerized simulation” vs. “computer simulation”, “electronic computers” vs. “microcomputers” vs. “personal computers”, and so on. The most important reason for this subcategory of inconsistencies is the lack of vocabulary control, i.e. many synonyms are simultaneously included in the thesaurus. It may have occurred at the beginning of the thesaurus’ construction. It may happen over time because of the delayed revisions of the vocabulary. Another reason is that indexers are not familiar with the subject areas; thereby they cannot accurately use these terms which are closely related or partially overlapping.

Subcategory 2: inconsistencies in using a broad term or a narrow term, for instance, “normal education” vs. “higher normal education”, “mental hygiene” vs. “mental health” vs. “mental health protection”, and so on. Usually, indexers should assign the most specific or co-extensive terms to a work. When such terms are not available in the thesaurus, broader terms or more general terms than the content of the work may be assigned. However, in practice, some indexers are more likely to assign broader terms when they are not sure which one is best-fitting due to their lack of subject knowledge. Sometimes, when indexers are not familiar with the thesaurus, they may assign broader terms as well, because it is much easier for them without digging deeply in the thesaurus to find an appropriate specific term. Moreover, the lack of detailed indexing policy brings about inconsistencies as well. For instance, whether or not should indexers assign a broad term after assigning a specific one? According to the data of this study, it happens very frequently that one indexer assigns a broad term after assigning a specific one, while the other indexer does not.

The most difficult situation to be handled by indexers is when there are no appropriate compound terms to describe complex topics. They usually need to coordinate two or more terms to represent the complex topics. In such case, in principle, they should pick up the most specific terms. But it is not always an easy task. For instance, when we want to represent the complex subject "teenagers' aesthetic education", and there is no such a specific descriptor in the thesaurus, we need to choose two terms and coordinate them to describe this complex subject. Should we choose "education of teenagers + aesthetic education" or "teenagers + aesthetic education"? According to the indexing polices which are applied by the two library systems, the former is preferred.

Subcategory 3: inconsistencies in direct entries or indirect entries, for instance, "software maintenance" vs. "software, maintenance", "adult higher education" vs. "adult education, higher education", and so forth. In some cases, a direct entry (or a specific term) exists in the vocabulary, but some indexers cannot find the right specific terms, then they use the coordination of two broader terms. The reason for this phenomenon is that these indexers are not familiar with the thesaurus. In some other cases, there are no direct entries (or appropriate specific terms) in the thesaurus for indexing, some indexers use appropriate specific terms extracted from free text to describe the documents (especially in computer science), while the others choose some broader terms from the thesaurus and coordinate them. The inconsistencies are resulted by the lack of a clear indexing policy. A strict rule should be set down so that indexers can clearly know under which situation a free-text term is allowed and under which situation it is not.

Subcategory 4: inconsistencies in free-text indexing or controlled indexing. It happens very frequently during indexing books of computer science. In computer science, some popular English terms are extensively used by authors in the literature and widely accepted by users. However, they are not included in the thesaurus, CALIS' policy allows indexers to use them at indexing, for example, "Windows", "FrontPage", "AutoCAD", and so on, while NLC has no such policy. The different policies result in inconsistencies.

In another case, in order to represent a book about UNIX operating systems, one indexer extracts an unauthorized term "UNIX operating system" from this book, while the other indexer uses a coordination of two authorized terms from the thesaurus "time-sharing system + operating system". It is the same as the second situation we have discussed in subcategory 3. To avoid this kind of inconsistencies, a detailed policy for free-text indexing is needed.

Subcategory 5: inconsistencies in descriptors or non-descriptors, for example, "interface (descriptor)" vs. "interface device (non-descriptor)". Obviously, some indexers are not familiar with the thesaurus or indexing rules, so that they make mistakes of this kind.

3.4.1.4..2 The inconsistencies in general terms and form terms

In relation to general terms, the inconsistencies can be further divided into two subcategories. One subcategory is that two indexers are not consistent in whether or not using a general term (the proportion is 89%). The other subcategory is that two indexers use different general terms (the proportion is 11%). Since general terms are usually too broad, it is very difficult for indexers to decide whether or not to use them to describe the documents in question. For example, to what kind of documents can be assigned the general terms "research", "methods", "knowledge", "basic knowledge", "primary knowledge", "theory", "basic theory", "principle", and so on? There are no clear definitions for these general terms in the thesaurus. Hence, what is the difference among "knowledge", "basic knowledge", and "primary knowledge" is a confusing question for indexers.

The inconsistencies of form terms can be further categorized in the same way as general terms. One subcategory is that two indexers are not consistent in whether or not using a form term (the proportion is 86%). The other subcategory is that two indexers use different form terms (the proportion is 14%). Form terms are usually utilised to describe what a document is, which is not the focus of the document, so indexers use them very freely at the indexing. On the other hand, some form terms are partially overlapping, and therefore confusing. For instance, "children's readings", "juvenile readings", "teenagers' readings", "young people's readings", "popular readings", "readings after school" and so on, are very confusing to indexers. Another example is: "reference materials for entering higher school", "reference materials for self learning", "teaching reference materials", "teaching reference books", "learning materials", "references" and "teaching materials". It is unreasonable that these terms are co-occurring in the thesaurus.

In fact, whether or not use form terms or general terms is still a controversial issue in the field of knowledge organization. On the one hand, form terms and general terms are too broad and can be applied to many documents, which makes them have too many postings. According to Karen Spark Jones, when there are too many postings under a term, this term is not discriminating any more. On the other hand, form terms and general terms can be very important for users at the time of filtering retrieving results, and meanwhile they are useful for collocation. Therefore, whether or not use form terms or general terms, to what extent use them, and how to use them appropriately still need to be researched.

3.4.1.4..3 The inconsistencies in geographic terms

The inconsistencies in using geographic terms can be further divided into the following four subcategories:

Subcategory 1: inconsistencies in using a full name or a shorten name (its proportion is 27%), for instance, "'Hebei" vs. "Hebei province" or "Beijing" vs. "Beijing city".

Subcategory 2: inconsistencies in whether or not using a geographic term (its proportion is 48%). For instance, indexers are very often not consistent in utilising the terms "China" and "World". Some indexers use them very frequently, while some others do not.

Subcategory 3: inconsistencies in using a narrower geographic term or a broader geographic term (its proportion is 18%), for instance, "Northeast of China" or "China".

Subcategory 4: inconsistencies in whether or not using a broader term after using a narrower term (its proportion is 7%). After assigning the name of a city, should indexers assign the name of the province, to which the city belongs?

The main reason for the first subcategory of inconsistencies is the lack of authority control. If the two catalogues use the same authority control files for the geographic names, this kind of inconsistencies can be avoided. For the other three subcategories of inconsistencies, we need to make a detailed indexing policy. For example, under what kind of situations should a geographic term be assigned? Should a direct entry or an indirect entry be assigned, namely, should the name of the place in question be directly assigned only or should the name of a larger geographic place be also assigned? Should the name of a location below the level of city be indexed? If the indirect entry is allowed, how many levels of geographic elements should be contained? We need to take into consideration all of these questions. If there is a clear and detailed indexing policy of using geographic names, and there are common authority control files, it is more likely for indexers to get high indexing consistency in using geographic terms.

3.4.2 Inconsistencies of notations

3.4.2.1 Different categories of inconsistencies and their proportions

Among the total of 3,307 pairs of records 1,439(43.51%) pairs are not exactly consistent in notations. The inconsistencies of notations can be categorized as follows:
Category 1: Difference on the top level
For example:
Title: Lords of the Harvest: Biotech, Big Money and the Future of Food
NLC: Index terms: Plant, genetic engineering, research
Classification: Q943.2 Plant genic engineering
CALIS: Index terms: Biotechnology, company, economy history, United States;
Record of actual event, United States
Classification: F471.267 Chemical industrial economy of United States
I712.5 Reportage of United States
This example shows that the two indexers have totally different subject analysis. The notations they have assigned to this book are different on the top level, i.e., they have chosen different main classes. The indexer in NLC considers that this book belongs to the subject area of botany, while the indexer in CALIS prefers that it belongs to the subject area of economy or literature.

Category 2: Difference on the coordinate level

For example:

Title: Learning theories for teachers

NLC: Index terms: learning theory, teaching psychology

Classification: G442 learning psychology

CALIS: Index terms: learning theory, teaching psychology

Classification: G422 teaching principles

This example is very interesting. The two indexers assign the same two terms, but different notations. They choose the same main class "G culture, science and education" and the same division "G4 education". However, the inconsistencies occur at the time of choosing a section. The indexer in NLC has chosen "G44 educational psychology", in contrast to the indexer in CALIS, who has chosen "G42 teaching theory". According to the abstract of this book, psychological aspects of various learning theories are discussed. Meanwhile, some suggestions are made to improve teaching efficiency. Therefore, we cannot say which one is right and which one is wrong, because both of the two aspects are included in this book.

Category 3: Difference on the hierarchical level

For example:

Title: Computer setting and optimizing

NLC: Index terms: Electronic computers, system performance, optimization

Notation: TP302 design and performance analysis

CALIS: Index terms: Electronic computers, system performance

Notation: TP302.7 performance analysis, function analysis

The indexer in CALIS has classified this book into a more detailed class than the indexer in NLC has. The relationship between the two notations is hierarchical. The following is the context of the two notations:

TP3 computing technology, computer technology

TP30 general issues

TP302 design and performance analysis

TP302.1 overall design, system design

TP302.2 logic design

TP302.4 graphics

TP302.7 performance analysis, function analysis

TP302.8 failure-tolerance technology

Category 4: Difference in the standard subdivision number

Title: When faster-harder-smarter is not enough: six steps for achieving what you want in a rapid-world

NLC: Index terms: success psychology, popular readings

Notation: B848.4 faith, will, behaviour

CALIS: Index terms: success psychology, popular readings

Notation: B848.4-49

In this example, the two indexers are totally consistent in index terms, but slightly different in notations. The indexer in CALIS has used the table of Standard Subdivisions, whereas the indexer in NLC has not. The standard subdivision "-49" stands for "popular readings".

Category 5: Difference in the indexing exhaustivity

For example:

Title: Inside macromedia Director 5 with Lingo for Windows

NLC: Index terms: multimedia, software tools

Notation: TP391 information processing

CALIS: Index terms: multimedia, software tools

Notation: TP391 information processing;

TP311.56 software tools, tool software

Here, the two indexers agree on all of the index terms and one of the notations. This example shows only one situation of this category. The example mentioned in category 1 shows the other situation, where all of the notations are different. Therefore, for this category, so long as the number of notations is different, we will say that the two indexers are inconsistent in indexing exhaustivity.

3.4.2.2 Distribution of different categories of inconsistencies

The relative frequency distribution of the five categories is shown in Figure 48. The fifth category represents the largest proportion, which is 28%. It means that the inconsistencies in indexing exhaustivity occur most frequently in comparison with the other four categories. How many notations should be assigned? What kind of aspects of the content should be represented? These questions are not easy for indexers. It is obvious that in most cases indexers are still used to assigning only one notation to a book (see Figure 18). However, more and more interdisciplinary fields are being created, and topics in documents are becoming more and more complicated. Describing them requires more than one notation. Thus, the conflict between the old custom (assigning only one notation) and the new requirement (assigning more than one notation) becomes more and more severe over time.

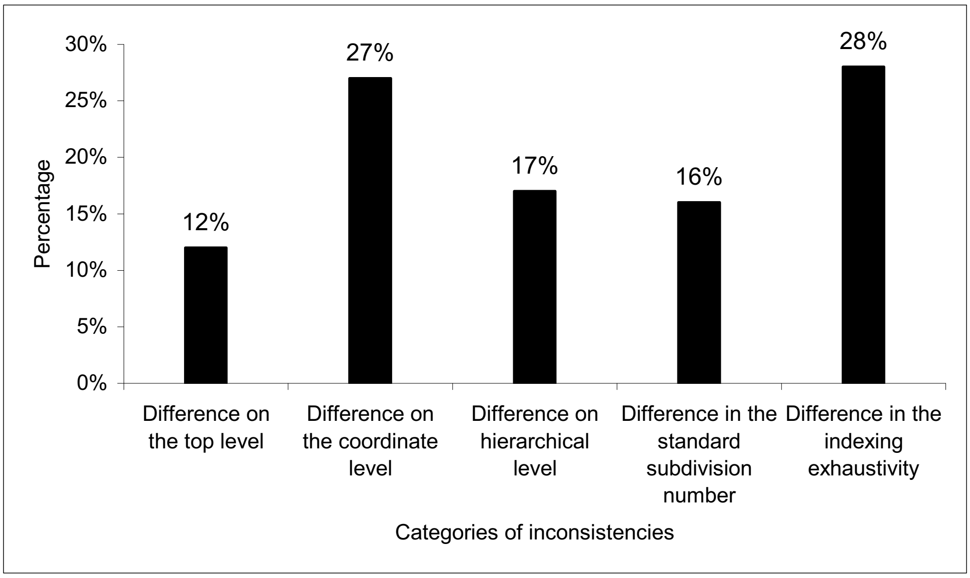

Figure 48: Relative frequency distribution of different categories of inconsistencies (notations)

The second category represents almost the same proportion as the fifth category. The difference on the coordinate level indicates that two indexers agree on selecting the main class and the division, but not on locating the most suitable placement within the division. Maybe the indexers are not familiar with the subject field, so that they cannot classify books into appropriate classes, or maybe the classification system itself is not good enough, in which the coordinate classes are too close to each other. That is to say, the difference among the coordinate classes is subtle, or the coordinate classes are not exclusive from each other.

In comparison with the second and fifth categories, the inconsistencies in the third category occur much less frequently. This kind of inconsistencies is resulted by different levels of indexing specificity. The general indexing policy applied by both of the library systems is that the most specific class number that represents the overall content of the work is assigned. If possible, the meaning of the number should be as specific as the content of the work being classified. When there is no co-extensive number, the next appropriate broader number is chosen. This policy is written on the paper. However, in practice, it is not always obeyed by indexers.

The relative frequency of the fourth category is similar to the third category. This sort of inconsistencies can be explained by the fact that the CLC falls between the enumerative and faceted classification. Most important subjects have ready-made class numbers that are enumerated in the CLC schedules. Nevertheless, many subjects are still not provided for, and these can be synthesized by the number-building process. On the one hand, the synthesis of class numbers can increase the capacity of the CLC to classify more and more subjects in ever greater detail. On the other hand, it can also cause indexing inconsistencies, because almost every number in the schedules can be further extended by notation from the table of Standard Subdivisions without specific instructions. It mainly depends on the indexer's decision.

The inconsistencies in choosing the main class occur least frequently (its proportion is 12%). Deciding to which subject field a book belongs doesn't require strong subject knowledge, but only some general knowledge. Nevertheless, judging the significance of a subject and then choosing an appropriate class requires the indexer to understand the message and purpose of the text and to be familiar with the practices and policies of the agency in which the indexer works. Sometimes, just because of following different policies or thinking from different viewpoints, indexers describe the same book in different ways. This does not imply that one of these indexing results is superior to the other; they are just different.

3.4.2.3 Comparison of the four subject areas

As mentioned above, the inconsistencies in choosing the main class occur least frequently. However, this category of inconsistencies occurs very frequently in psychology. In comparison with the mean value of this category (12%), its proportion (22%) is relatively high in psychology (see Figure 49). If we examine the records of this subject field closely, we can find that it may be due to the intersection between psychology and other subject areas, for instance, sociology, education, etc. Usually, "see also" references, which link the related classes, are utilised to resolve the conflict of linearity of classification scheme and interdisciplinarity of subjects. However, "see also" references can confuse the indexer at the time of choosing appropriate classes, or even make the indexer's decision more difficult, when the annotations are not clear. In fact, because of the limitation of the scheme's volume, the annotations can not be very detailed. Sometimes there is only a short sentence under a class, which indicates its related class (or classes). But there is no description about classifying principles, i.e. what kinds of documents should be classified into this class, and what kinds of documents should be classified into the other. Then indexers may interpret the annotations in different ways, which can result in inconsistencies.

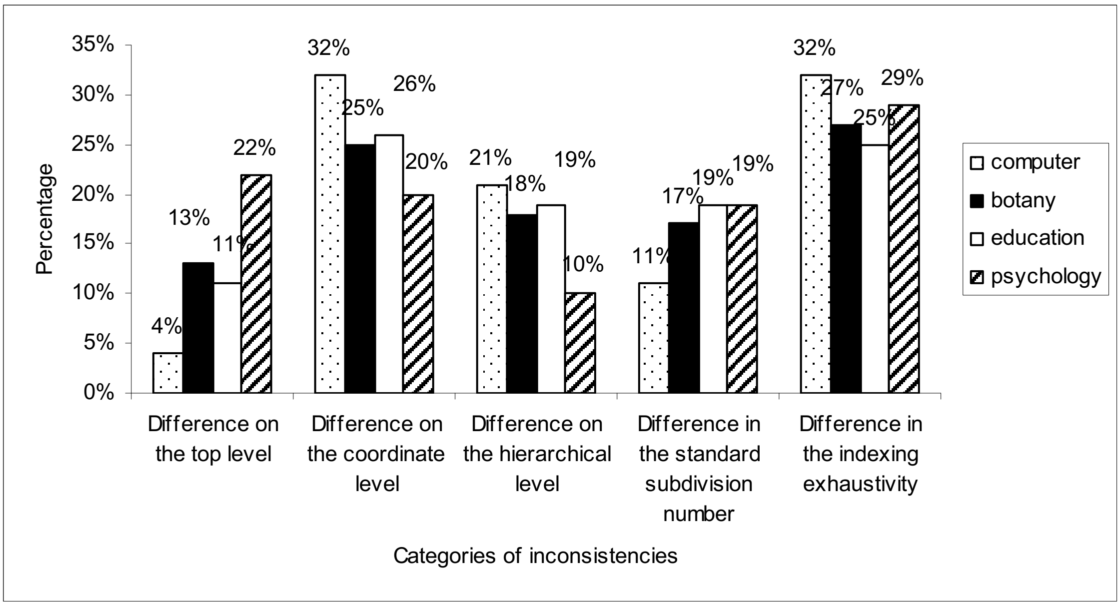

Figure 49: Comparison of different categories of inconsistencies in notations among the four subject areas

Another reason for this category of inconsistencies is that many books in psychology in relation to the sample are popular readings rather than scientific monographs. It is known that bibliographic classifications are usually based on the classification of disciplines. The definitions of classes are usually scientific and strict. Therefore, it is somewhat difficult for indexers to choose an appropriate discipline (a main class) to describe a popular reading book, because sometimes they cannot decide to which discipline this book belongs. This problem can be exemplified as follows:

Title: How to get promotion, how to start the undertaking and how to make a pile: 66 skills

NLC: Index terms: Success psychology- popular readings

Notation: B848.4 Interests and attitudes

CALIS: Index terms: career, creativity, enterprise management

Notation: C96 talent studies

This book is a popular reading. It deals with the topic of how getting success in career. To which discipline does it belong? It is not clear.

In botany, the reasons for this category of inconsistencies are a little different. By analysing the sample, we can find that the class "S agriculture" is the one which is frequently mixed with the class "Q94 botany" by the indexers. On the one hand, it is because of the intersection between agriculture and botany. However, the most important reason is that the two systems adopt different indexing policies. CALIS is an academic library system. Many university libraries are as its members, among which there are also some agriculture universities. If there is a book dealing with plants, the indexers in these agriculture universities are inclined to classify this book into "S agriculture" rather than "Q94 botany". For instance,

Title: The genus fusarium

NLC: Index terms: fusarium

Notation: Q949.32 epiphyte (Q949.3 fungus)

CALIS: Index terms: fusarium

Notation: S432.4 contagious disease (S432 plant diseases and their prevention)

Objectively, this book should be classified into Q949.32. But, as special libraries, they prefer the class "S432.4". We cannot say which choice is wrong and which is right, because they have different target readers. Classifying a book into the most useful class is a basic principle for the indexers.

It is surprising that the proportion of this category of inconsistencies in computer science is very low (4%). Although computer science is a relatively new discipline and computer technology has been used in many other fields, there is a unified policy of indexing documents of computer science, that is, a work treating the application of computer technology to other disciplines is classed with the disciplines being acted on. This policy brings high consistency.

For computer science, the biggest problem is the inconsistencies on the coordinate level. Its proportion is 32%. On the one hand, the scheme of computer science is too old. Some classes are obsolete or out of date. And the whole structure of this scheme is unreasonable. The bad knowledge system causes the low consistency, because the indexers cannot find appropriate classes when indexing. The following is the outline of the scheme of computer science:

TP3 Computing technology and computer technology
- TP3-0 Computer theories and methods
  - TP3-05 The relationship between computer and other disciplines
- TP30 General issues
  - TP301 Theories and methods
  - TP302 Design and performance analysis
  - TP303 Overall structure, system structure
  - TP304 Materials
  - TP305 Manufactures, assembling, reassembling
  - TP306 Adjustment, testing, verification
  - TP307 Repair, maintenance
  - TP308 Computer rooms
  - TP309 Security and encryption
- TP31 Computer software
  - TP311 Programming, software engineering
  - TP312 Programming languages, algorithmic languages
  - TP313 Assemblers
  - TP314 Compliers, interpreters
  - TP315 Management programs, management systems
  - TP316 Operating systems
  - TP317 Software packages (general purpose application programs)
  - TP319 Specific application programs
- TP32 General calculators and computers
  - TP321 Non-electronic computers
  - TP322 Analytic computers
  - TP323 Electronic calculators
- TP33 Electronic digital computers

TP331 Basic circuits
TP332 Processor/ CPU
TP333 Storage
TP334 Peripherals
TP335 Information conversion and its devices
TP336 Bus, channels
TP337 Simulator
TP338 Various electronic digital computers

TP34 Electronic analogue computers
TP342 Processor/controller
TP343 Storage
TP344 Input/output
TP346 Function generator
TP347 Time-lapse apparatus
TP348 Various electronic analogue computers

TP35 Mixed electronic computers
TP352 Digital-analogue computers
TP353 Analogue-digital computers

TP36 Microcomputers
TP368 Various microcomputers

TP37 Multimedia technology and multimedia computers

TP38 Other computers
TP381 Laser computers
TP382 Beam computers
TP383 Superconductive computers
TP384 Molecular computers
TP387 Fifth generation computers
TP389.1 Artificial neural nets computers

TP39 Computer application
TP391 Information processing
TP392 Various specific databases
TP393 Computer networks
TP399 Application in other fields

In order to prove the fact that this structure is unreasonable, I count the number of documents under every class in computer science in the NLC. The data in the following table depicts the distribution of the documents under different classes.

Table 15: The number of postings under the classes of computer science

classes	No. of documents all years	No. of documents from 1988 to 2004
tp3	3,447	3,392
tp30	1,177	1,067
tp31	10,393	10,097
tp32	33	6
tp33	740	348
tp34	12	2
tp35	3	1
tp36	3,054	2,784
tp37	254	250
tp38	58	51
tp39	19,139	18,875

We can see from Table 15 that there are too many documents under the classes “TP31Computer software” and “TP39 Computer application”, while there are almost no documents under the classes “TP34 Electronic analogue computers” and “TP35 Mixed electronic computers”. It is really unusual that the documents under these classes distribute so unevenly.

Obviously, the unreasonable classification scheme makes the indexing more difficult. However, on the other hand, as the indexers lack subject knowledge, the situation becomes worse. Although the terminology in computer science is not as abstract as in social science and humanities, it changes very quickly. The indexers should be qualified to do the job. They should know about the computer technology and should have the basic knowledge of this subject field. Otherwise, the indexing quality cannot be improved even if a good classification scheme is available.

In comparison with computer science, the same problem occurs in education, but for different reasons. In education, the proportion of the second category of inconsistencies, i.e. the inconsistencies on the coordinate level, is 26%. Theoretically speaking, it is mainly because that there are many classes in the classification scheme in this field (552 classes). Therefore, the granularity of the classes is small. In another words, the distances between the coordinate classes is small, while the similarity between them is large. Thus, it is relatively difficult to distinguish them when indexing. The following is the outline of the scheme of education:

G4 Education
- G40 Pedagogy
- G41 Political and moral education
- G42 Teaching theories
- G43 Audio-visual education
- G44 Educational psychology
- G45 Teachers and students
- G459 School and family, school and society
- G46 Education administration
- G47 School management
- G48 School buildings and materials management
- G51 Educational system of the world
- G52 Educational system of China
- G53/7 Educational system of other countries
- G61 Preschool education
- G62 Elementary education
- G63 Secondary education
- G64 Higher education
- G65 Normal education
- G71 Vocational education
- G72 Adult education
- G74 Overseas Chinese and emigrant education
- G75 Minority education
- G76 Special education
- G77 Social education
- G78 Family education
- G79 Self-learning

In practice, we find that this kind of inconsistencies occurs mostly under the following situations:

Firstly, the indexers do not pay attention to the indention of a class, which indicates the hierarchical position of this class in the overall classification. For instance:

Title: The effects of schooling

NLC: Index terms: Education- effects- research

Notation: G40 pedagogy

CALIS: Index terms: School education- research- education evaluation

Notation: G449.7 Evaluation methods and criterion

The context of the class "G449.7" is as follows:

G4 Education
  G44 Educational psychology
    G449 Educational psychology testing and evaluation
      G449.7 Evaluation methods and criterion

Obviously, the indexer in CALIS has made a mistake, because he has not noticed that the broader class of "G449.7" is "educational psychology", which does not fit in the content of this book.

Secondly, the indexers mix the general teaching theory with the special teaching theory of various levels (for instance, elementary school, middle school, high school, etc.) or of various subjects (for instance, mathematics, physics, chemistry, English, etc.). For example:

Title: The basis of research learning theory

NLC: Index terms: teaching research, elementary and middle school

Notation: G632.0 teaching research and reformation

CALIS: Index terms: teaching methods, research; scientific research, capability cultivating.

Notation: G424.1 teaching methods

The following is the contexts of the two classes:

G63 secondary education
  G632 teaching theories, teaching methods
    G632.0 teaching research and reformation

G42 teaching theories
  G424 teaching methods and teaching organizing
    G424.1 teaching methods

This book deals with the general learning theory rather than the learning theory of secondary education. Thus, the indexer should be familiar with the division of the classification scheme. Usually, bibliographical classification groups like concepts together and related concepts near each other in a manner that places the whole before its parts. The indexer should be aware of the difference between the general aspects and the specific aspects of the scheme.

Thirdly, the indexers often mix the "teaching references" with the "teaching materials, textbooks and learning references". For example:
Title: the complete reference of physical experiments for middle school
NLC: Index terms: physics, experiments, middle school, handbook
        Notation: G634.7 learning references of physics for secondary education
CALIS: Index terms: physics, middle school, teaching references, experiments
        Notation: G633.73 teaching references of physics for secondary education
Theoretically, "teaching references" are for teachers, while "learning references" are for students. However, in practice, a book can be used by both of the two groups. From this example we can see that this book can be used either by a teacher or a student. It indicates that the difference between the two classes sometimes is not obvious. In order to avoid this kind of confusion, these two classes should be revised.

Concerning the third category of inconsistencies, i.e. the inconsistencies on the hierarchical level, computer science stands out again (its proportion is 21%), and education is in the second position (its proportion is 19%). For computer science, the reasons are almost the same as those we have mentioned above. The unreasonable structure of the classification scheme makes the indexers confused with the knowledge system and makes it hard to classify. Further more, computer science and education share some common reasons for the inconsistencies on the hierarchical level.

Firstly, in the schedules there are some special auxiliaries, which express aspects that are recurrent. These special auxiliaries provide the means of representing complex subjects in minute classification. Meanwhile, there are also some subdivisions by analogy from the schedules themselves. Whether or not adding a number from these special auxiliaries and subdivisions by analogy to the base number is not obligatory. It is more up to the indexers, which can easily result in the inconsistencies on the hierarchical level.

Secondly, the inconsistencies in indexing specificity occur frequently when indexing comprehensive works. In both of the two schedules with regard to the two subject fields, there are some classes named "theories". For example,

Title: cooperation with children's nature: a young German's education dream in China.

NLC: Index terms: children education, quality education

Notation: G610

CALIS: Index terms: children education, quality education, research

Notation: G61

The following is the context of these two classes:

G61 preschool education

G610 preschool education theories

G611 moral education

G612 teaching theories and teaching methods

G613 teaching methods and teaching materials of various subjects

G614 toys and teaching aids

G615 staff of preschool education

G616 kindergarten and family

G617 kindergarten management

G618 various types of kindergartens

G619 general situation of preschool education of the other countries in the world

From the abstract in the records, we can know that the author of this book is a young German, who has stayed in China for six years. In this book, on the basis of his own experiences and his observations, he discusses about some education issues in China, and he suggests that the quality education of children should be adopted to resolve these problems. Can we say that the subject focus of this book is “preschool education theories”? Or is it just a comprehensive book about “preschool education”? What kind of books can be regarded as dealing with “preschool education theories”? Is there distinct difference between a comprehensive book and a book dealing with “theories”?

Thirdly, there is also a “specific entry” policy for classifying, i.e. the most specific class number that represents the overall content of the work should be assigned. Ideally, the meaning of the number should be co-extensive with, that is, as specific as, the content of the work being classified. However, in many cases, there is no co-extensive number (either pre-existing or newly established). Usually, the next appropriate broader number is chosen. This is the rule, which is printed on the paper. In practice, the indexers choose a broader number not only when there is no co-extensive number, but also when they can not make sure which one is the appropriate one. In order not to deviate far from the right number, they just adopt a conservative method, i.e. choosing a broader number.

As mentioned above, the Chinese Library Classification falls between enumerative and faceted classification. Most important subjects have ready-made class numbers that are enumerated in the CLC schedules. Nevertheless, many subjects are still not provided for, and these can be synthesized by the number-building process. Number building is the process of constructing a number by adding notation from the tables or other parts of the schedules to a base number. Broadly speaking, there are three methods of number building in CLC:

- adding the number from the common auxiliary tables
- adding the number from the special auxiliary tables
- adding the number through the subdivision by analogy from the schedules themselves

Almost every number in the schedules can be further extended by notation either from one or more of the auxiliary tables or from the schedules themselves. Sometimes a number can be added to the base number in the schedules without specific instructions (e.g. the number from Standard Subdivisions), but in most cases a number building should be according to instructions found under a particular entry.

We have discussed as above that adding a number from the special auxiliaries and the subdivisions by analogy to the base number can easily cause inconsistencies on the hierarchical level. Although this number building process should be conducted according to instructions, in many cases, it is not obligatory. It is even worse when adding the number from Standard Subdivisions to the base number without any specific instructions.

Referring to the category of inconsistencies in the Standard Subdivision number, education has the same proportion (19%) as psychology does. It is mainly because that the documents in these two subject fields are more likely to deal with some general concepts or some form aspects. And these general concepts and form aspects are usually only minor aspects of a document. Meanwhile, whether or not adding the number from Standard Subdivisions to the base number is an arbitrary decision.

With regard to the fifth category of inconsistencies, i.e. the inconsistencies in exhaustivity, computer has the highest proportion (32%), and then psychology (29%) follows. For computer science, the main reason is the unreasonable structure of the classification scheme. When the indexers cannot find an appropriated class in the scheme, they will assign two or more class numbers to bring out the subjects of the book in hand. For psychology, the main reason is its intersection with the other subject fields. When the content of the work covers more than one topic or aspect, and there is no available number that covers all topics or aspects treated in the work, a choice often must be made between two or more numbers, each of which covers part or parts of the content of the work. If the indexers insist on choosing only one class number, the inconsistency on the top level can occur. If one indexer decides to choose more than one number, while the other still insists on choosing only one number, the inconsistency in exhaustivity can occur.

For the other two subject fields, this category of inconsistencies occurs also very frequently. In fact, it becomes a big problem for all the subject fields. On the one hand, there is a tendency that the topics in books become more and more complicated or there are more and more books with compound topics. On the other hand, the bibliographic classification has the limitations imposed by its linearity. The conflict between the complexity of the topics and the linearity of the classification scheme makes the indexing process more difficult. Although several solutions to solve this problem have been brought out, new problem arises. One of the solutions is to define rules such as, "class by the emphasis of the work" or "choose a broader number" or "choose the number of the subject being influenced or being acted on", and so forth. Another one is to allow the indexers to assign more than one number. The result is that the classifier sometimes must make rather arbitrary decisions. It may result in the inconsistencies on different levels or the inconsistencies in exhaustivity.

3.4.2.4 Reasons for different categories of inconsistencies

3.4.2.4.1 Category 1: Difference on the top level

There are many factors that result in this category of inconsistencies. We can investigate them from the following four viewpoints:

Firstly, from the viewpoint of indexers, due to their different knowledge backgrounds and their different experiences with subject indexing, they can make different subject analyses for the same book, therefore, choose different main classes.

Secondly, from the viewpoint of indexing policies, different library systems may have different intended readers; therefore they may have different indexing policies. As discussed above, an agriculture university library is inclined to class a book about plants with the main class "agriculture" rather than "botany". Thus, different indexing policies can cause inconsistencies in choosing the main classes.

Thirdly, from the viewpoint of the classification scheme, on the one hand, there are some "see also" references among different disciplines which are originally used to link the related classes and to help users search documents from different perspectives. However, when the annotations under these related classes are not clear, or not complete, the indexers may interpret the annotations in different ways, which can result in inconsistencies. On the other hand, the classification scheme is a discipline-based classification system and multidisciplinary topics or marginalized fields are not best handled in it. The classes do not fulfil the traditional requirement of being mutually exclusive and jointly exhaustive. The overlaps between classes can easily cause the inconsistencies.

3.4.2.4.2 Category 2: Difference on the coordinate level

Just as discussed above, this kind of inconsistencies occurs very frequently. It can be analyzed from two aspects.

The first one is the structure of the classification scheme. In this classification scheme, some classes, especially in computer science, are out-of-date. When indexers cannot find an appropriate class, they may select a related class haphazardly. Thus, different indexers may have different selections. In some cases, the illogical relationships between coordinated classes can confuse indexers too. Traditionally, the coordinate classes should be exclusive from each other, and the whole array should be exhaustive of the contents of the class. This is best achieved by using only one characteristic of division for any one hierarchical level. Nevertheless, it is sometimes necessary to use not only logical divisions but also other, sometimes, overlapping, characteristics of division based on what kinds of documents exist in the respective fields. Such overlapping characteristics of division result in confusion and make it hard to make classification decisions. Moreover, when the granularity of the classes is too small, i.e. the distances between the coordinate classes is too small, the difference of the coordinate classes is so subtle that the indexers cannot distinguish between them, especially when there are no clear and complete annotations. In some other cases, there are two alternative classes in the classification scheme, which are used not by force of the classification system, but are decided by the library’s indexing policy or by the indexer. So, this flexibility of using the alternative classes will to some extent cause the inconsistencies.

The second one is the qualification of the indexers. The indexers are usually not the subject experts. Sometimes they cannot understand the tiny and very special differences between the coordinate classes. Thus, they cannot translate their subject analysis into indexing language correctly even if they make a right decision at the stage of subject analysis. In addition, sometimes the indexers are not familiar with the classification system or the indexing rules. In practice, they make mistakes mainly because they only pay attention to the name of a class but fail to notice its context in the overall hierarchy.

3.4.2.4.3 Category 3: Difference on hierarchical level

Actually, inconsistencies on hierarchical level are inconsistencies in indexing specificity. It is a complicated issue. There are some reasons like the following:

First, different library systems may adopt different policies of specificity due to their own intended reader groups or the characteristics of their collections. For instance, in a public library system, a learning reference book about English for middle school pupils would be represented with a class comprising the aspects of "learning reference book", "English (as a subject)", and "middle school". It is because there are too many such books in a public library system and there are many readers who need to distinguish a learning reference book from a dictionary, or a book for middle school from a book for high school, and so on. However, in an academic library system, firstly there are not many such books; secondly, the readers are usually academicians. They don't have needs for such detailed information. Therefore, the indexers in an academic library system may class the same book with a class comprising only the aspect of "English (as a subject)".

Second, the unreasonable structure of the classification scheme can cause inconsistencies in this category. Because when there are no appropriate or specific classes in the scheme for a book, some indexers may class it with a broader class number, whereas other indexers may class it with a coordinate class number which has close meaning. The two different ways in which they index the same book result in inconsistency in specificity.

In addition, this category of inconsistencies occurs very frequently during the number-building process. Adding a number from the special auxiliaries and the subdivisions by analogy to the base number can easily cause the inconsistencies on the hierarchical level. On the one hand, this process is not obligatory. It depends on the content of the books, the indexing policies of the library systems, or the indexers' arbitrary decisions.

On the other hand, it is a complex process. When indexers are not familiar with the classification system or the indexing rules, they may easily make mistakes.

The last point is that some indexers are not familiar with the classification system or they don't totally understand the exact subject a book deals with, so they cannot judge correctly which class will be the best choice for the book, and then they adopt a very conservative method "upload to the broader class". This may also cause inconsistencies.

3.4.2.4.4 Category 4: Difference in the standard subdivision numbers

The table of Standard Subdivisions is a miscellany of recurring concepts, forms, and etc. It is utilized to add dimensions to the classification schedule, thereby to some extent counteracting the limitation of its linearity. It is convenient to represent the content of a document from different dimensions. However, in a paper-based classed catalogue, the standard subdivision number cannot be used alone when users search the catalogue, but can only be used together with the base number. Its primary function is to help users filter the searching results. Generally, the standard subdivision number can only represent the minor subject of a document. Thus, it is more likely to be ignored by indexers for its unimportance. In fact, in computerized files, classification potentially permits searching by other than the first element in a combined notation, and searching by forms other than subject (such as: within a span of numbers search only textbooks). Thus, in OPACs the standard subdivision numbers become more and more important.

Since the standard subdivision number can only represent the minor subject of a document, the biggest problem for the indexers is to whether or not add a standard subdivision number. Thus, to some extent the inconsistencies are inevitable because of the flexibility of using the table of Standard Subdivisions and the inherent characteristic of its reflecting minor aspects of documents. Another reason is that the break-down of this table is a little haphazard, because more than two characteristics of the division are used at one time. It makes the relationship between classes illogical and confusing, and as a result the indexers misunderstand the meanings of some classes. This misunderstanding may lead to inconsistencies in adding standard subdivision numbers.

3.4.2.4.5 Category 5: Difference in the indexing exhaustivity

Bibliographic classification systems, especially enumerative classifications, structure the world of knowledge with disciplines. Meanwhile, the original purpose of library classification is to arrange books on the shelves of a library in a sequence according to their subjects, so that users can find on the shelves like items that they might wish to use at the same time. And because a book can only have one position on the shelf, there is a rule that only one notation can be assigned to a book.

However, newly developed topics straddle more than one discipline and are marginalized in the classification scheme's structure. At the same time, there are more and more books dealing with multidisciplinary or interdisciplinary subjects. It makes it difficult for the indexers to class those books with only one notation. On the other hand, the topics in books become increasingly specific and complex. It encourages the libraries to make analytic indexing along with the traditional emphasis on overall content.

In connection with OPACs classification systems can not only be used to arrange books, but also be used to retrieve books in the database. That is, the notation that places a document into an intellectual class structure does not have to indicate its physical location. In order to make it clear, we call the notation used for bibliographic access and retrieval "subject notation", and the notation used for the physical location "call number". In OPACs, the subject notation does not need to be the same as the call number. The differentiating between the subject notation and the call number permits assigning more than one bibliographic retrieval notation to a document.

Although assigning more than one notation can counteract the limitations of linearity of the classification system and to some extent resolve the problem of the interdisciplinary and the compound topics, it brings out another problem, which is the inconsistency in the indexing exhaustivity. On the one hand, the idea of assigning more than one notation is not yet commonly accepted by all the libraries. On the other hand, some indexers are used to assigning only one notation. They can only change it gradually. Moreover, there are no well-defined rules for assigning multiple notations. Therefore, it more depends on the indexer's arbitrary decision. More detailed indexing policies are seriously needed to change this chaotic situation.

4 Discussion

4.1 Introduction

The problem of indexing consistency has been addressed in numerous studies in the past. Markey (1984) summarized 25 consistency studies and concluded that "the interpretability of interindexer consistency scores is both frustrating and bewildering in the light of their variability". According to her summary of the 25 studies, the consistency scores range from 4% to 82%. Nevertheless, many researchers believe that indexing consistency is inevitably low. However, the data of this study demonstrate relatively high indexing consistency. An in-depth discussion of the big variability of indexing consistency is carried out in Section 4.2, while the reasons for the relatively high consistency of this study are discussed thoroughly in Section 4.3. Further more, some contradictions between this study and previous studies in relation to the factors affecting indexing consistency are discussed in Section 4.4.

4.2 Discussion of the big variability among the results of the previous studies

Studies of indexing consistency reported in the literature have shown that the consistency values vary a great deal. Thus, many researchers cautioned us not to compare those studies and their consistency scores, for instance, Markey (1984) claimed: "In fact, the studies and their scores ought to be considered separately and not compared." Nonetheless, some researchers still believe that in general the indexing consistency is low. For example, Bates (1986) summarized the previous studies by saying: "no matter what measure of consistency was used, rates of inconsistency in indexing were found to be high. Two indexers well trained in an indexing system (interindexer consistency) would frequently index a given document differently, and even the same indexer (intraindexer consistency) would use different terms at different times on the same document. For example a typical result is found in a study by Jacoby and Slamecka, reported in Stevens: The interindexer consistency was found to be 20% and the intraindexer consistency 50%".

Is it really true that 20% and 50% are typical figures for the consistency studies? Let us see what Jacoby and Slamecka themselves said. The following statement claimed by them was cited by Mann (1997) as the argument against the unreasonable generalization of different consistency studies: "The resulting percentages of indexer reliability do not then necessarily reflect the average performance of experienced indexers, and – since the conditions of the experiment differed from those of everyday indexing practice of the contractor—they do not form a basis for an evaluation of the indexing system used. The conclusions derived here apply only to the conditions of this experiment; and it should not be assumed, without further evidence, that they hold also for situations in which one or more of the parameters have been changed (i.e., the system of indexing, the body of documents, etc.)" Obviously, Jacoby and Slamecka did not claim that 20% and 50% were typical or reflected the average performance of experienced indexers. They clearly pointed out that their conclusions applied only to the conditions of their experiment. Therefore, any sort of generalization of the consistency studies should be made very carefully.

Why is there such a big variability among the results of the former studies? It can be attributed to several factors. About thirty years ago, Leonard (1977) already realized the big variability of the results of the consistency studies and mentioned some reasons for this variability: "these studies use several different measures of indexer consistency. They vary considerably in research methodology and in the treatment of variables for the indexers, the documents selected, and the controls imposed on the indexing process. Not surprisingly, they also vary considerably in the level of indexer consistency reported." Markey (1984) also mentioned two factors, namely, the formulas used to measure consistency and the definitions of exact matches. However, they only mentioned those reasons rather than discussed them in detail. Actually, so far no one has discussed the reasons for the big variability in depth. In the following texts the reasons for the big variability are analyzed from six aspects.

4.2.1 Different research methods (experimental vs. observational)

The using of different research methods is one of the factors contributing to the incompatibility of the previous studies. In summary, two categories of research methods have been used in the earlier studies, namely, experimental and observational. For instance, the following studies are experimental: Harwitz, 1969; Zunde & Dexter, 1969; Leonard, 1975; Markey, 1984; Iivonen, 1990; Chu & O'Brien, 1993; David et al., 1995; Bertrand & Cellier, 1995; Collantes, 1995; Stubbs et al., 1999; Shoham & Kedar, 2001; Saarti, 2002. While the studies conducted by Funk et al. (1983), Middleton (1985), Chan (1989), Booth (1990), Sievert & Andreas (1991), Tonta (1991), Reich & Biever (1991), Svenonius & McGarry (1993), Leininger (2000), Subrahmanyam (2006), etc. are observational.

In the experimental researches, usually an experiment is designed to simulate an indexing environment. The researchers choose some indexers and some documents. The indexers should index these documents in certain time. After that the researchers compare the indexing results and calculate the indexing consistency. This method is to some extent reasonable, because the researchers can manipulate the variables, i.e., they can test how two variables co-vary by setting the other variables unchanged. For example, the researchers can let two indexers index one book twice, one time with a controlled vocabulary, the other time without any controlled vocabulary but with their own languages. By doing this, the researchers can investigate whether there is any difference in indexing consistency between controlled indexing and uncontrolled indexing. Here "indexing method (controlled vs. uncontrolled)" is an active independent variable. According to Leech et al. (2005), "an active independent variable is a necessary but not sufficient condition to make cause and effect conclusions; the clearest causal conclusions can be drawn when participants are assigned randomly to conditions that are manipulated by the experimenter." It indicates that if we want to make cause and effect conclusions, there should be an active independent variable which can be manipulated. This is the advantage of the experimental studies, because only experimental studies have an active independent variable.

However, several limitations are apparent in the experimental studies. One limitation is that the samples of this kind of experiments are usually small. Regarding the sample size we will discuss in detail in the next subsection. Another limitation is that there is very likely to be Hawthorne effect in this kind of experiments. That is, "when such studies are intentionally conducted, indexers are frequently aware that they are part of a study, which may improve performance" (Leininger, 2000). Funk et al. (1983) realized this disadvantage of experimental studies as well. In their article, they commented: "because of each study's experimental design, each indexer was aware that he was being 'tested' on his consistency. This awareness could have produced indexing consistency percentages vastly different from those seen in normal working conditions."

In order to overcome these shortcomings of experimental studies, more and more researchers are inclined to draw the samples directly from the real cataloguing environment of a large database. These studies are observational or non-experimental studies, in which there are no active independent variables but only attribute independent variables. The attribute independent variables cannot be manipulated. For instance, after the researchers choose a database to draw samples, the variable "indexing method (controlled vs. uncontrolled)" is not an active independent variable any more, rather an attribute independent variable, whose value is not changed during the study.

Comparing with experimental researches, observational researches overcome almost all the shortcomings of the experimental researches. In principle, there is no limitation of small sample size and no Hawthore effect any more. Thus, the results are more objective, more reliable and closer to the reality. However, they have their own problems. Since in observational researches we cannot manipulate independent variables, we can only make indirect inferential conclusions rather than direct causal ones. In most cases, there are many independent variables which can simultaneously influence the dependent variable(s).

In conclusion, the using of these two different research methods in part resulted in the variability of the consistency values in the earlier studies.

4.2.2 Different sample size

Sample size is another factor which contributes to the big variability of the consistency scores in the previous studies. In general, the larger the sample size, the closer the results approach to the reality. If sample size is too small, the experiment will lack precision to provide reliable answers to the questions it is investigating. If sample size is too large, time and resources will be wasted, often for minimal gain. With respect to the consistency studies, many of them have very small sample size, which is shown in Table 16.

Table 16: The sample sizes of the previous studies

research	sample size	research	sample size
Painter (1963)	321	Rodgers(in Markey, 1984)	20
Hooper (1965)	25	Zunde & Dexter (1969)	29
Bryant, King & Terragno (in Leonard, 1977)	201	Funk et al. (1983)	760
Gehring (in Leonard, 1977)	10	Markey (1984)	100
Scheffler (1968)	35	Chan (1989)	100
Kyle (in Leonard, 1977)	246	Iivonen (1990)	10
Mullison et al. (1969)	50	Booth (1990)	28
Tagliacozzo&Kochen (1970)	1123	Sievert & Andreas (1991)	71
Korotkin & Oliver (in Leonard, 1977)	30	Tonta (1991)	82
Stevens (1965)	20	Reich (1991)	236
Knapp (1944)	219	Svenonius (1993)	100
Borko (1964)	997	Chu & O'Brien (1993)	3
Slamecka & Jacoby (1963)	75	Bertrand (1995)	8
Bates (1977)	30	David (1995)	2
Lancaster (1968)	16	Collantes (1995)	40

Fried & Prevel (1966)	45	Stubbs et al. (1999)	1
Hurwitz (1969)	31	Leininger (2000)	60
Harris et al. (1966)	3	Shoham (2001)	50
Tarr & Borko (1974)	15	Saarti (2002)	5
MacMillan (1961)	171	Olson & Wolfram (2005)	1
Leonard (1975)	60	Medelyan & Witten (2006)	10
Lilley (1954)	6		

It is indicated by Table 16 that 72% of these studies have a sample size smaller than 100 items. 21% of the studies have a sample size smaller than 500 items. Only 7% of the studies have a sample size larger than 500 items. The reason for the small sample sizes is the considerable cost of processing record twice or the difficulty of finding the duplication of the records in one system. However, the samples are too small for researchers to draw valid conclusions from, because such small samples might not be representative of the overall consistency of indexing in the data base (Funk et al., 1983). Therefore, the findings of these studies cannot be generalized.

4.2.3 Different methods of measuring consistency

"Another variable that affects indexing consistency is the consistency measure used in the evaluation. Studies reported in the literature employed a variety of methods and different formulae to calculate indexing consistency values." (Tonta, 1991)

In the second chapter we have discussed the different formulae which were used to calculate the consistency in the earlier studies. Using of different formulae inevitably results in different consistency scores. For the same set of data, the consistency scores calculated with Rolling's formula are usually higher than those calculated with Hooper's formula. And many studies even didn't use any formula to calculate the consistency.

Moreover, different criteria have been used to determine "exact matches" in earlier studies, which brought about the variability of indexing consistency as well. Actually, some researchers mentioned it in their articles. For example, Reich and Biever (1991) commented: "This wide range in the consistency figures can be attributed, at least in part, to the criteria used to determine matches. Some investigators considered a match to have occurred if there was agreement on the indexed concepts, while other investigators required agreement on terminology." And Iivonen (1990) proved that there were clear differences in consistency figures depending on whether it was calculated on the basis of terms or on the basis of concepts or on the basis of aspects.

Even when the calculation of matches is on the basis of terminology, the results can be very different, when different kinds of comparison unit are used. The comparison unit can be an index term or a subject heading string. When the indexing language involved is a thesaurus, the unit is usually an index term; when the indexing language involved is a subject heading system, the unit is usually a subject heading string. And the consistency scores calculated based on index terms are usually higher than on subject heading strings. In the second chapter, the difference of the two kinds of comparison unit has been discussed in detail.

4.2.4 Different types of documents (monographs/articles/other kinds of documents)

In the previous consistency studies, different types of documents were involved. Some studies investigated books, some studies dealt with journal articles, some studies observed non-print materials like artworks and pictures, and some studies related to general documents. Since different types of documents have their own characteristics, they are indexed based on different indexing policies.

For example, there is a big difference between the indexing of books and the indexing of journal articles. In general, when indexing books, terms are assigned to represent the overall content of the book, i.e., the content of an entire book or serial, rather than that of individual chapters or articles. Lancaster (1998) described the situation of book indexing as follows: "the subject matter of books is represented at a very general and superficial level (an average of fewer than two subject heading/ subheading combinations per item as reported by O'Neill and Aluri, 1981)." In comparison with the indexing of books, the indexing of journal articles is more analytical. Journal articles are usually indexed more exhaustively, i.e. with more terms. Moreover, the terms assigned to a book are more likely to be collective or generic, while those assigned to a journal article are more specific. The indexing of images or art works is obviously a little different. They need to be indexed at an exhaustive level, because it is difficult to identify the aboutness of them (Enser, 1995; Markey, 1984).

Actually, in the previous consistency studies, there were some evidence showing the big difference in exhaustivity and specificity among different types of documents. The exhaustivity and specificity may to some extent affect indexing consistency. Moreover, the difficulty of getting consistency varies among different types of documents. Thereby, the results of the former studies involved different types of documents can not be compared with each other.

4.2.5 Different types of indexers (experienced vs. non experienced/ subject specialists vs. non subject specialists)

Indexers as actors of the indexing process play a significant role on the indexing results. Iivonen (1990) found out that those indexers using UDC in their job used more qualified terms of time and place than other tested persons. Lancaster (1998) assumed that two indexers with very similar backgrounds (education, experience, interests) would be more likely to agree on what should be indexed than two indexers with widely differing backgrounds.

Actually, in many of the previous experiments, the subjects (indexers) involved were the students without real indexing experience or even the users of the libraries. The consistency observed based on these novice indexers cannot stand for the real situation of the experienced indexers in the libraries. Mann (1997) argued that the indexing consistency of the novice indexers (users or students) using uncontrolled indexing method could not be used as the basis to challenge the consistency or the worth of the work done by trained professional cataloguers using controlled vocabularies. And many studies (Gehring (1964), Jacoby & Slamecka (1963), Leonard (1975), Shoham & Kedar (2001), Saarti (2002)) have shown that the consistency among experienced indexers is usually higher than that among non-experienced indexers, although there is a study not supporting this statement (Fried &

Prevel, 1966). However, there was no apparent evidence showing a significant correlation between consistency and educational background or subject speciality (Jacoby & Slamecka (1963), Leonard (1975), Korotkin & Oliver (1964), Bertrand & Cellier (1995)).

4.2.6 Different types of indexing methods (controlled indexing vs. uncontrolled indexing)

Whether use a controlled vocabulary during the indexing process is another factor that can influence indexing consistency. Markey (1984) summarized 25 studies and enumerated 5 studies which supported the statement that "interindexer consistency improves when indexers choose index terms from a controlled vocabulary". However, relationship between vocabulary control and indexer consistency is not as simple as it seems to be at the first sight. Lancaster (1998) elaborated on the complicated relationship between vocabulary control and indexing consistency. He concluded that "a controlled vocabulary should improve consistency of indexing in the long run but it can only be applied consistently by experienced indexers knowledgeable in the subject matter and fully familiar with the terms" and that "controlled vocabulary, by reducing the amount of choice, has a beneficial effect on consistency of indexing over a large group of documents."

Although the relationship between consistency and vocabulary control is complex, we must admit that the consistency of controlled indexing to some extent is different from that of free-text indexing. In the former studies, some experiments involved controlled indexing, some involved uncontrolled indexing, and some compared the difference in consistency between controlled and uncontrolled indexing. Therefore, we cannot simply compare the consistency scores recorded in the former studies while ignoring the indexing methods involved in those studies.

4.2.7 Conclusions

There may be some other reasons for the variability of the consistency scores in the earlier studies. It appears that consistency values depend on a number of factors under which the indexing is performed. We should be very careful when comparing the figures of those studies. Only when two experiments have the same experimental conditions, the results are comparable. Actually, on the topic of indexing consistency, very few studies have the same experimental conditions. Thus, it is almost impossible to compare the consistency scores. Nevertheless, we can discuss the different conclusions in those studies, for example, in which ways a factor affects indexing consistency.

Any review of this topic should pay more attention to the different conditions of every study. It is unreasonable to compare the scores only and then make inferential conclusions while ignoring the different methodologies of those studies. In fact, meta-analysis of this topic is urgently demanded. "Meta-analysis is a statistical manoeuvre for combining results across studies to reach conclusions, entails the mathematical synthesis of results from independent studies" (Ankem, 2005). Since it is not the focus of this study, we will not discuss it here in detail. Ankem (2005) offered a detailed discussion about meta-analysis and introduced some methods of conducting meta-analysis in library and information science.

4.3 Discussion of the reasons for the relatively high consistency of this study

After summarizing the earlier studies, we can now discuss the results of this study. The consistency scores in this study seem relatively high. The average consistency of index terms is 64.21%, and the average consistency of class numbers is 61.58%. Both of the consistency figures are calculated according to Hooper's formula. Why are the scores in this study relatively high? Many factors contribute to the high consistency scores.

4.3.1 Comparison unit

First, it is due to the comparison unit used when calculating indexing consistency. This issue has already been discussed in Subsection 2.5.1. The results are very different, when different kinds of comparison unit are used. If we recalculate the indexing consistency of the same sample based on term strings, we can find that the overall consistency is 47.4%, whereas it is 64.21% when the comparison unit is a term.

We may say that the value 47.4% is a typical value in consistency studies, and it confirms Bate's conclusion that "no matter what measure of consistency was used, however, rates of consistency in indexing were found to be low, frequently less than 50 percent" (Bate, 1986). Low consistency makes some people loose their confidence in subject cataloguing. Gregor & Mandel (1991) commented in their article that "it is an impossible and unrealistic goal to attempt to collocate all books on a certain subject under a single subject term… Indexers have long understood that only a reasonable degree of interindexer consistency can ever be achieved." Is it really such frustrating? Which of the two figures (47.4% vs. 64.21%) is more representative?

Before we answer this question, we can take a look at online catalogue users' subject searching behavior at first, because the interaction between subject searching and subject indexing should be taken into account when we evaluate the quality of subject indexing. Lancaster (1998) defined good indexing as "indexing that allows items to be retrieved from a database in searches in which they are useful responses and prevents them from being retrieved when they are not". This implies that the quality of indexing should be examined from the point of view of users' subject searching.

In fact, users' subject searching behavior has been extensively discussed in the last decades. Yu and Young (2004) believe that the popularity of the Web has influenced users' mental models and thus their expectations and behavior when using a Web-based OPAC interface. On the one hand, there is a tendency that subject searching is being replaced by keyword searching in online catalogues (Larson, 1991). On the other hand, users incline to search Web-based OPAC using single terms rather than complicated term strings. Pulis and Ludy (1988) reported that user searches were typically single words. According to Jansen and Pooch (2001), the majority of searchers on both OPACs and Web search engines use approximately two terms in a query. Although Hildreth (1997) found that more than two-thirds of keyword searches included two or more words, it was shown that 42 percent of these multiple-word searches resulted in zero hit.

Now, we can go back to the question. We can say that 64.21% is more representative of the indexing consistency in the context of this study, because it is more meaningful if we calculate indexing consistency based on terms rather than term strings. Since users infrequently retrieve exact matches on long term strings, it is not necessary to calculate indexing consistency based on term strings. As long as one term in a term string matches, users can retrieve the books which are indexed with this term in both of the library systems. From the data of this study we can see that indexing consistency of index terms is not as low as many researchers have thought it would be. And there is still room for improvement in it. The idea of giving up subject indexing because of its nature of subjectivity is wrong.

4.3.2 Experienced indexers and controlled vocabularies

The second point is that this study is an observational research. The collected data are the real bibliographic records from two library systems' bibliographic databases. The sample size is relatively large (3,307 books, i.e., 6,614 records). And the subjects involved in this research are professional library cataloguers. They index books obeying rigid indexing rules and using controlled indexing languages. If we compare the results of this study with the results of the former experimental researches, we can say that it is not surprising that we get relatively higher consistency scores, because most of the experimental researches involve non-experienced indexers conducting natural language indexing. As Mann (1997) argued, "consistency cannot be achieved by people who rely on natural language key words as indexing terms, in the absence of vocabulary control mechanisms to lead them to proper headings". In fact, there are some pieces of evidence showing that the consistency among experienced indexers using controlled vocabularies is higher than that among novice indexers with free-text indexing (see Subsection 4.2.5 and 4.2.6).

4.3.3 Online access

Another reason for the relatively high consistency scores is the ease of consulting their own catalogue or other online catalogues during indexing process in the networked environment. When indexers index a book in hand, they can firstly search in their own or other online catalogues, just like conferring with other professional colleagues. If the same book or a similar book is found, they can directly accept other indexers' subject description, which may lead to high consistency, or make some small modifications. O'Neill and McCain (1995) reported that 80% of the DDC class numbers assigned by the LC were accepted by the local library and another 10% were accepted with minor revision.

In fact, a recent enhancement of *WebDewey* makes it much easier for indexers to check their own catalogue while they are attempting to develop the correct Dewey Decimal Classification (DDC) number for an item they are cataloguing. Gonzalez (2005) explained the function of consulting one's own library's catalogue during indexing. She said: "When classifying a resource, you may need to check your own library's bibliographic database to determine what your past practices have been and to compare the item in your hand with other items already catalogued in your collection. Seeing what you've already classed in a certain number that you are considering for a new item, and whether the works in that number are similar to the work you have in your hand, helps you determine a 'fit' ".

The web-based LCC, *Classification Web,* has the similar features, i.e. indexers can consult their own library's catalogue or other catalogues during classifying. Moreover, *Classification Web* allows indexers to write and save personalized notes and search these notes. In addition, a *Classification Web* e-mail discussion list exists. Indexers can discuss any ideas or problems there. All these functions are beneficial to indexing consistency. A current research conducted by Subrahmanyam (2006) indicates an extremely high level of consistency in the light of LCC numbers. He found that a title is likely to have the same class number in more than 85% of the library systems holding it. It is understandable that such high degree occurs in his study, because situation is changing in the networked environment. Many improvements have been made on classifications, which make it more possible to achieve high indexing consistency.

4.3.4 Classification number and descriptor correlations

A further point is the use of the Chinese Classified Thesaurus during indexing. It is an integration of the Chinese Library Classification and the Chinese Thesaurus, that is to say, a two-way corresponding list between class numbers and descriptors (Zhang et al., 1996). A system for merging classification and controlled subject vocabulary – mapping against each other – has great potential for improving indexing consistency.

When indexers choose one class number, they can get suggestions of correspondent descriptors at the same time, or vice versa. Theoretically, it can save indexers' time and can to some extent improve indexing consistency.

Actually, it can be justified by the result which has been discussed in Subsection 3.3.5. The data indicate that there is a strong correlation between consistency of index terms and consistency of class numbers. When the consistency of index terms is high, the consistency of class numbers is very likely to be high, vice versa. An explanation for this interesting phenomenon is that indexers use the Chinese Classified Thesaurus as a reference tool during indexing. That is to say, when they choose the same classification number, they are very likely to choose the same descriptors, because these descriptors are suggested in the corresponding list. The number of the suggested descriptors in the corresponding list is much smaller than that of the descriptors in the whole thesaurus. Thus, it is much easier for indexers to choose a best-fit descriptor and more likely to be consistent with other indexers.

Nowadays, it is a tendency in this field to provide indexers correlations between classification numbers and subject headings or descriptors. Both *WebDewey* and *Classification Web* have provided such enhancement by displaying correlations between classification numbers and subject headings.

4.4 Discussion of the contradictions between this study and previous studies in relation to the factors affecting indexing consistency

As mentioned above, there are many factors which can affect indexing consistency. Zunde and Dexter (1969) listed 25 factors influencing indexing consistency. Many other researchers discussed some of the factors as well. However, concerning the ways in which these factors influence indexing consistency, the researchers are not consistent with each other. Markey (1984) enumerated the studies supporting one statement and those not supporting it. And she underlined that the findings of interindexer consistency experiments are inconsistent particularly in the evaluation of subject specialists, indexing aids, and indexer experience. In fact, there are some contradictions between this study and the previous studies as well. It involves the factors exhaustivity, specificity and length of item. In this section we discuss these contradictions in detail.

4.4.1 Exhaustivity

Since exhaustivity is an important factor which affects retrieval effectiveness, it has been investigated in indexing consistency studies as well. A few researchers have attempted to explore the relationship between exhaustivity and indexing consistency. 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), Reich & Biever (1991), Sievert (1991), Shoham (2001). However, this is not always the case, because the findings of the studies conducted by Harris et al. (1966), Leonard (1975) 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 caused by different levels of exhaustivity or by different number of descriptors which were identical to terms in title.

Sievert (1991) compared the consistency of main headings (50.39%) with that of subheadings (47.89%) and the consistency of descriptors (47.27%) with that of identifiers (32.83%). He believed that the differences were caused 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. Lancaster (1998) assumed that "one would expect more agreement on what are the main topics of the document than on which of the minor topics are worth inclusion." Funk et al. (1983) proved that the difference in consistency scores between central concepts and non-central concepts was statistically significant. Iivonen's (1990) experiment also showed that indexers were more likely to agree on central aspects of a work than on peripheral themes. Thereby, it is not surprising that the consistency of main headings was higher than that of subheadings in Sievert's study. 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). It is understandable that the consistency of descriptors (controlled indexing) is higher than that of identifiers (uncontrolled indexing). Therefore, it is not convincing to say that higher exhaustivity results in lower consistency based on the findings of Sievert's study.

Actually, if we want to speculate whether there is a correlation between indexing consistency and exhaustivity, we should control other variables, so that we can observe whether consistency co-varies with the number of descriptors. We cannot determine whether there is a correlation between indexing consistency and exhaustivity, when some other variables are involved, for instance, types of index terms, types of indexing methods, etc.

Shoham and Kedar (2001) have made a considerable improvement over previous studies. They directly calculated the correlation between indexing consistency and the 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 the number of terms was not substantially strong.

As mentioned above, some researchers took a contrary position. Harris et al. (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.

Upon examination of previous studies, we may conclude that the relationship between indexing consistency and exhaustivity is complex. It is not necessarily true that assigning more index terms results in lower consistency, because it depends partly on whether two indexers assign equal number of index terms to the same book, i.e., whether they have identical indexing depth. The data of this study indicates that when two indexers have identical indexing depth, indexing consistency is substantially high. On the contrary, when they have different indexing depths, consistency is significantly low. If we examine previous studies in depth, we can find some evidence for it.

Reich and Biever (1991) realized that different indexing depths could result in low consistency and they said that "the difference in the number of terms assigned to each document obviously contributes to the observed low index match rate of 27 percent." They presented a scatter diagram of the number of NAL and CAB terms per document, which showed relatively few instances of identical indexing depth (see Figure 50). They described this diagram as follows: "the diagram depicts a scattering distribution rather than the clustering along with a straight line that would be suggested by the overall averages."

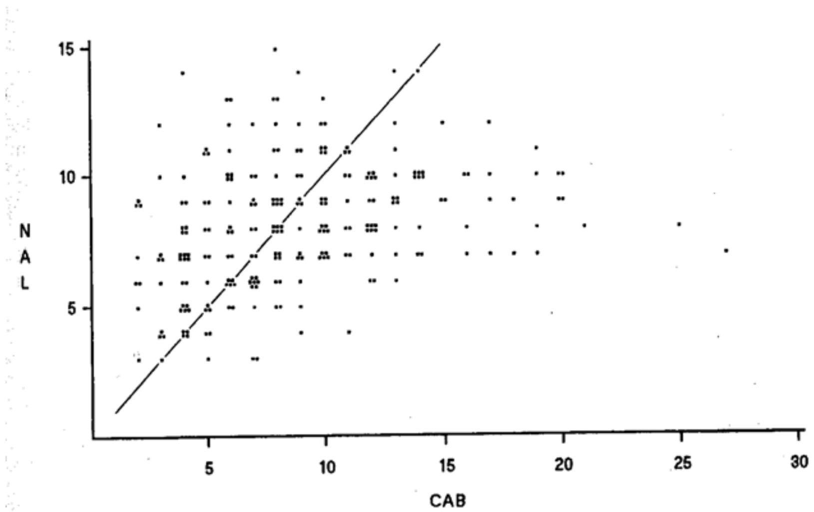

Figure 50: Different indexing depths in Reich & Biever's study

In comparison with their study, the data in this study show a clustering distribution along with diagonal cells, which indicates that the exhaustivity of NLC is equal to that of CALIS (see Table 17). This is one of the important reasons for the high consistency in this study.

Table 17: Identical indexing depths in this study

nlc \ calis	1 term	2 term	3 term	4 term	5 term	6 term	7 term	8 term	9 term
1 term	258	73	24	10	2	2	0	0	0
2 term	105	788	189	65	25	5	0	0	0
3 term	70	168	645	133	33	15	2	1	0
4 term	8	49	79	292	44	12	5	1	0
5 term	3	11	13	26	60	12	5	2	1
6 term	0	4	10	4	8	23	3	1	1
7 term	0	0	0	0	1	1	3	3	1
8 term	0	0	2	0	1	0	0	2	0
9 term	0	0	0	0	0	1	0	0	1

Legend: the numbers in this table stand for number of books

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 catalogers assigned 282 subject headings for 82 items while BL catalogers assigned 127. In other words, on the average, LC assigned 3.44 subject headings per title. The same average was 1.55 for BL catalogers. The difference in indexing depth between the two catalogues was significant.

Stubbs' experiment (1999) also provided 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 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.

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. Actually, as we have discussed in Subsection 3.3.2, it is inevitable with the use of the existing formulae that two indexers are more likely to be consistent with each other when they have identical level of exhaustivity.

4.4.2 Specificity

In previous consistency studies, a few researchers investigated the relationship between consistency and specificity. Lancaster (1998) summarized these studies (Tinker, 1966, 1968; Lancaster, 1968; Lancaster, 1964; Sinnett, 1964; Mullison et al., 1969; Reich & Biever, 1991) and made the following conclusion: "the larger the vocabulary, the more specific it is likely to be; and, the greater its specificity the more difficult it will be to use consistently." It seems to contradict the result of this study. However, if we thoroughly inspect these studies, we can find that this issue has been discussed based on different understandings and definitions of specificity.

For instance, in the context of Lancaster's statement, "specificity" means specificity of a vocabulary rather than specificity of an entry or a term. Specificity of a vocabulary refers to the level of detail of the terminology in the whole vocabulary. It is determined when the vocabulary is constructed, whereas specificity of an entry or a term is in relation to indexing. A highly specific vocabulary is able to express the subject in great depth and detail.

These studies, which were cited by Lancaster to support his position, did not directly discuss the relationship between consistency and the size and specificity of vocabulary. They can be divided into two groups. The first group (Tinker, 1966, 1968; Lancaster, 1968; Reich & Biever, 1991) investigated different categories of indexing consistency and found that the lowest consistency occurred in the category of the combinations of main headings and subheadings. The second group (Lancaster, 1964; Sinnett, 1964; Mullison et al., 1969) dealt with the role indicators' effect in reducing consistency.

Firstly, we analyze the first group of studies. Although combinations of main headings and subheadings are usually used to describe documents more specifically, they are not equivalent to specificity of vocabulary. When we construct a subject heading system, the inclusion of combinations of headings and subdivisions can increase the specificity of this subject heading system. But we cannot increase the specificity of the vocabulary by using such kinds of combinations at indexing time. During indexing, if a co-extensive heading does not exist in the vocabulary, indexers will establish a combination of main heading and subheading(s), so that they can exactly describe the content of the work being catalogued. In this sense, at indexing time the combinations of main headings and subheadings are used to achieve high level of indexing coextensivity rather than to increase specificity of vocabulary. Thus, we cannot use these studies to support the position that greater specificity of vocabulary leads to lower consistency. When we talk about combinations of main headings and subheadings leading to low consistency, we cannot conclude that greater specificity of vocabulary causes lower consistency.

Actually, the fact that the combinations of main headings and subheadings lead to low consistency can be interpreted in a different way. We can say that indexers are more likely to be consistent when choosing terms directly from the vocabulary than making a combination of multiple terms. Slamecka and Jacoby (1963) and Lancaster (1998) explained the phenomena through a distinction between "prescriptive" and "suggestive" vocabularies. Slamecka and Jacoby believed that "inter-indexer consistency improves significantly with the use of prescriptive indexing aids containing a minimum of variable semantic relationships among terms. The use of indexing aids which enlarge the indexer's semantic freedom of term choice is detrimental to indexing reliability." Lancaster gave a good example to explain it: "Indeed, it seems probable that the greatest consistency would be achieved in the assignment of those terms that might be preprinted on an indexing form or displayed online (as in the case of the 'checktags' of the National Library of Medicine) to remind an indexer that they must be used whenever applicable."

Concerning the second group of studies, I'd like to cite a brief definition of "role indicator" from ODLIS (Online Dictionary for Library and Information Science) at first. Role indicator is defined as follows: "In indexing, a code used to indicate the syntactic relationship between two or more index terms (subject headings or descriptors) assigned to a document to facilitate retrieval by subject." According to this definition, we can see that role indicators can only function as indicators of the syntactic rather than the semantic relationship between two or more index terms. Syntax treats the arrangement and relative positions of terms in a sentence or statement. Using role indicators in indexing can avoid ambiguity at the retrieval stage rather than increase specificity of vocabulary.

To summarize, these studies cited by Lancaster to support his position seem to be irrelevant to the issue of the relationship between consistency and specificity of vocabulary. Thus, there is a lack of evidence to conclude that greater specificity of vocabulary causes lower consistency.

However, since Lancaster discussed specificity at the vocabulary level, whereas this study at the term level, we cannot compare this research with Lancaster's. Hence, an observation is made based on the data of this study to speculate whether there is a correlation between consistency and the size and specificity of vocabulary. The result is tabulated in Table 18 and shows no evidence of larger vocabularies resulting in lower consistency. For instance, in the light of the number of terms, the vocabulary of botany is the largest, but the consistency of botany is also the highest. On the other hand, in terms of the number of classes, the vocabulary of education is the largest; however, the lowest consistency occurs in computer science rather than in education.

Table 18: The relationship between the size and specificity of vocabularies and consistency

Subject area	Number of classes	Consistency	Number of terms	Consistency
Botany	187	65.89%	1514	70.87%
Computer	196	54.59%	492	56.43%
Psychology	66	65.31%	no data	67.37%
Education	552	60.99%	no data	62.84%

Note: The number of terms in psychology and education can not be counted because the full Chinese Thesaurus is not available in Germany, where this study is conducted.

Theoretically, neither Lancaster's conclusion nor the result of this study is wrong, because large size and great specificity of vocabularies can be detrimental as well as beneficial to indexing consistency.

On the one hand, increasing specificity of vocabulary in a subject area would increase granularity. In other words, the more terms are contained in a controlled vocabulary, the smaller distance is among terms, and that is, the larger similarity is among them. Then, it is more difficult for indexers to distinguish them, which causes difficulty in achieving high consistency. In this sense, increasing the size and specificity of a controlled vocabulary is detrimental to indexing consistency.

On the other hand, if the size of a controlled vocabulary is larger, i.e. more terms being contained, then the possibility of finding appropriate terms is greater. If indexers can find best-fitting terms in the vocabulary, they don't need to figure out a combination of multiple terms. In fact, indexers are more likely to be consistent when choosing terms directly from the vocabulary than making a combination of multiple terms. In this sense, indexing consistency benefits from the larger size and greater specificity of the vocabulary.

It seems to be a paradox that increasing the size and specificity of vocabularies is detrimental as well as beneficial to indexing consistency. Nevertheless, we must face the reality that the size and specificity of vocabularies have two sides of effect on indexing consistency. This is why we should optimize an appropriate level of specificity when constructing a thesaurus.

From the analyses of this subsection and of Subsection 3.3.3 we may conclude that indexing consistency is not always negatively correlated with specificity. With respect to the specificity at term level, there is no linear relationship between consistency and term frequency. In other words, greater specificity does not necessarily bring about lower consistency. Concerning the specificity at vocabulary level, the indexing consistency depends more on the characteristic of subject areas and the structure of the vocabulary than on the size and specificity of the vocabulary.

4.4.3 Length of item indexed

Whether length of item indexed influences indexing consistency is still a controversial issue so far. Some scholars have investigated it and made totally contradictory conclusions. Firstly, Zunde and Dexter (1969) explored whether a document's readability influenced interindexer consistency and found that the readability of document was not shown to strongly influence indexing consistency. In their article, they defined "readability" as a function of "$RE = 206.835 - 0.846W_1 - 1.015 S_1$", where W_1 denotes the average number of syllables per 100 words, S_1 denotes the average sentence length in words. Thus, the "readability" does not actually equal to length of item indexed, because it considers only sentence length and word length rather than length of the whole item indexed. In the light of their definition of "readability", they did not find significant improvement in consistency among indexers of documents with high reading ease.

Later, both Markey (1984) and Lancaster (1998) made a summary of this issue in their reviews of consistency studies, and they drew the same conclusion that short items were indexed more consistently than long items. Markey said in her article: "The complexity of documents affects interindexer consistency. Four experiments provided evidence that interindexer consistency increases as documents decrease in length". She enumerated some studies to support this conclusion (Rogers (1961); Fried and Prevel (1966); Leonard (1975)). Lancaster had the same belief, although he expressed it in a different way. In his book, he stated: "the length of the item indexed should affect consistency: the shorter the item, the fewer the terms that might plausibly apply." To support his statement, he cited several studies as well. Except for the studies cited by Markey, he mentioned another three studies (Harris et al. (1966); Horký (in Lancaster, 1998); Tell (1969)).

It is very confusing that Markey and Lancaster have made two entirely opposite interpretations of Tell's study. According to Markey, "Tell (1969) reported that there was not an improvement in interindexer consistency when indexers proceeded from indexing documents according to their abstracts to using their full text" Whereas, Lancaster claimed that "Tell (1969) discovered that consistency when indexing from the full text of articles was lower than when indexing from titles or abstracts." If we look at Tell's original article, we can find a paragraph in the "conclusion" section as follows: "Compared with an indexing standard, the indexing made by the journal editors and reviewed by specialists, indexers are more consistent when indexing from titles or abstracts than from full text. It has been shown that indexers have difficulties in behaving consistently when using full text documents compared with a standard." From this paragraph it can be seen that Markey has misinterpreted Tell's study.

All the misinterpretations and contradictory conclusions make us confused. What should we believe now? Should we believe that indexing consistency increases as documents decrease in length? Or should we not? In order to answer these questions, we can make a distinction among the studies which deal with this issue and divide them into three groups.

The first group, of which the study of Zunde and Dexter's (1969) is a good example, explores whether a document's readability influences indexing consistency. As mentioned above, "readability" is not identical to "length of item". It concentrates only on sentence length and word length. The conclusion of the first group is that readability of document does not strongly influence indexing consistency.

The second group of studies tries to find out how a document representation in the form of a title, an abstract, and a full text affects the choice of indexing terms and to relate that to indexing consistency. For instance, Harris et al. (1966) compared the indexing of questions (short text) and that of journal articles, and found that the consistency in indexing questions was higher than in indexing journal articles. Tell (1969) discovered that consistency was lower when indexing from the full text of articles than when indexing from titles or abstracts. Slameck & Jacoby (1965) also claimed that "Indexing from titles or abstracts …, is more consistent than 'random' indexing from the entire document." From the findings of this group of studies we can draw a conclusion that consistency is higher when indexing from compressed short texts than when indexing from full texts.

The third group of studies intents to investigate whether cataloguers index long articles or monographs more consistently than index short ones. For example, Funk et al. (1983) compared the indexing of journal articles with different pages and claimed that: “The articles in our sample were from one to forty-five pages long, with a mean of 6.3 pages. A Pearson correlation showed no statistically significant correlation between length and consistency for any of the nine categories of consistency.” The current study conducted by Shoham & Kedar (2001) showed that “as the extent of the item rose (up to 600 pages) the level of consistency fell, and then rose again for extremely long books (over 600 pages).”

Since the indexing of monographs with different pages has been observed in this study as well, a comparison can be made between this study and Shoham & Kedar’s. In order to make this comparison clear, the data of the two studies are tabulated (see Table 19). Firstly, if we conduct t-tests on their data, we can find that the consistency (28.10%) of the books in the range of 400 to 500 pages is significantly lower than the books in other ranges of length ($p<.05$). Except for this value, there is no significant difference among the other consistency values which are shown in the third column of Table 19.

Table 19: Comparison of the data of this study and Shoham & Kedar’s data in the light of relationship of consistency and length of item indexed

Length of item (no. of pages)	Average Indexing Consistency of this study (Hooper’s formula)	Average Indexing Consistency of Shoham & Kedar’s study (Hooper’s formula)
From 100 to 200 p.	66.84%	34.50%
From 200 to 300 p.	62.68%	32.04%
From 300 to 400 p.	63.09%	34.03%
From 400 to 500 p.	67.08%	28.10%
From 500 to 600 p.	64.67%	29.56%
Over 600 p.	63.368%	34.98%

If we plot the data on a graph (Figure 51), we can see that indexing consistency does not vary much with length of items in both studies, which is indicated by the small slope values of the two regression lines (-0.0025, -0.0031). The consistency values are obviously higher in this study than in theirs. The reasons for the high consistency of this study can be found in Section 4.3.

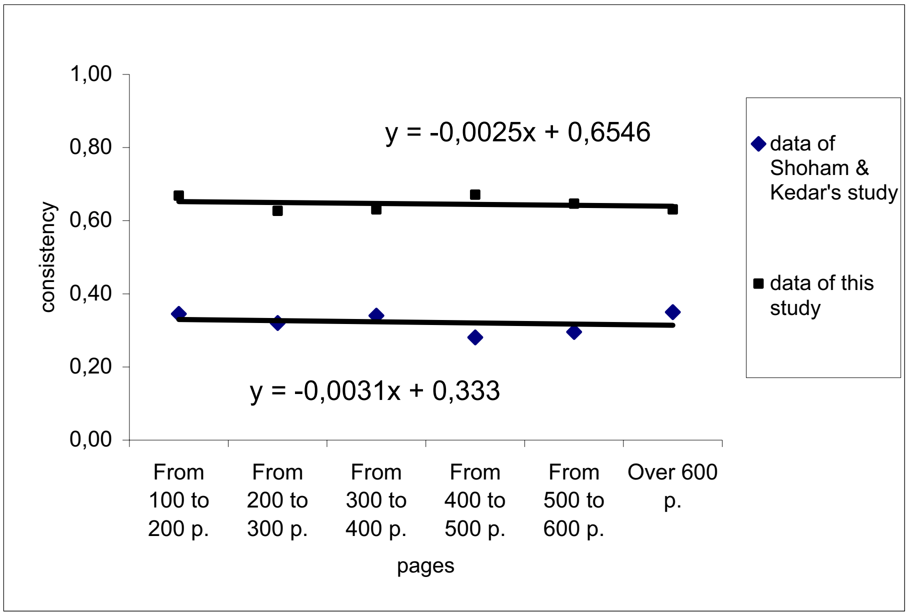

Figure 51: Comparison of this study and Shoham & Kedar's results in the light of relationship of consistency and length of item indexed

Therefore, it can be dangerous to make a generalization from their findings that indexing consistency declines with length of items up to a point and then it increases again. In fact, from this multiple line graph we can see that the indexing consistency is almost constant with length of items, which implies that length of items has almost no effect on indexing consistency. From this group of studies we can conclude that there is no distinct correlation between length and consistency.

Now we can go back to these questions. Should we believe that indexing consistency increases as documents decrease in length? Or should we not? In different cases, the answers are different. Just as analyzed above, the consistency in indexing from compressed short texts (titles, abstracts, etc.) is higher than in indexing from full texts, whereas readability of documents and length of articles or books do not strongly influence indexing consistency.

Why can the answers be so different? In the first case, the compressed short texts, e.g. titles, abstracts, etc., are condensed descriptions of full texts. The central topic(s) or the central aspect(s) of a topic is (are) more likely to stand out in titles or abstracts than in full texts. Thus, it is easier for cataloguers to index from titles or abstracts consistently. When a full text is used as indexing source, indexers firstly need to analyze the central topic(s) or the central aspect(s) of a topic, and then compress the whole full text into some index terms in relation to the central topic(s) or the central aspect(s) of a topic. In this progress, indexers should differentiate central topics from peripheral ones, and sometimes it may be very confusing for them. In this sense, it is not surprising that indexing consistency is lower when indexing from full texts.

Fried and Prevel (1966) claimed: "As the quantity of information in an article decreased, indexing accuracy and consistency generally increased." However, longer articles or books do not necessarily contain more quantity of or more complicated information. For example, a dictionary focusing on a topic may be very long. When indexing a dictionary of such kind, indexers usually apply only two terms, one topical term and one form term. In such case, indexing consistency is more likely to be high. On the contrary, shorter articles or books may deal with very complex or multiple topics, which may result in lower consistency. Thereby, when we talk about the relationship between indexing consistency and length of items, we should discuss it in different cases. If we mix them up, a variety of confusing conclusions may occur.

In brief, the consistency is higher when indexing from compressed short texts (titles, abstracts, etc.) than when indexing from full texts, whereas readability of documents and length of articles or books do not strongly influence indexing consistency.

5 Conclusion

5.1 Summary of the study

The data of this study demonstrate relatively high indexing consistency, given the study's stringent condition imposed for "exact match". According to Hooper's formula, the average consistency for index terms is 64.2% and for class numbers 61.6%. According to Rolling's formula, for index terms it is 70.7%, for class numbers 63.4%. Although the scores are not impressively high, they are much higher than the commonly accepted figure "less than 50%" (Bate, 1986). The high consistency figures imply that situation is changing in the networked environment. In an OPAC environment, conventional bibliographic classification systems and thesauri are utilised not only as organizing tools, but also as online searching tools. On the one hand, many improvements have been made on classifications and thesauri, which make it more possible to achieve high indexing consistency. On the other hand, the compatibility of different indexing tools is also beneficial to indexing consistency.

In relation to the factors affecting indexing consistency, the following conclusions can be made based on the data of the study:

1. Consistency tends to be higher when more terms are totally assigned to a book by two indexers.

2. Two indexers are more likely to be consistent with each other when they have identical level of exhaustivity. Moreover, the smaller difference in indexing exhaustivity, the greater probability they get to achieve high consistency. It is inevitable with the use of the existing formulae.

3. The more terms two indexers totally assign to a document, the smaller are the differences in the consistency scores, no matter how many terms they commonly assign. It is inevitable with the use of the existing formulae.

4. Term specificity has no influence on indexing consistency. Whether or not high consistency can be achieved depends more on the characteristic of subject areas than on the size and specificity of the vocabulary.

5. Length of articles or books does not influence indexing consistency.

6. There is a positive correlation between the consistency of index terms and the consistency of class numbers. That is to say, when the consistency of index terms is high, the consistency of class numbers is very likely to be high as well, and vice versa.

7. Generally speaking, lower indexing consistency can be observed in humanities and social science, the so called soft science. However, low consistency can also occur in technology and natural science (hard science), when a discipline develops so quickly that the revision of indexing languages cannot keep pace with its development.

8. The dramatic change or development of a discipline like computer science may lead to the unstable terminology, and thereby may result in lower consistency in the indexing of documents in this subject area.

9. The unreasonable or obsolete structure of vocabularies can cause low consistency.

10. The consistency declined by one percent per year over the last decades, because on the one hand many new concepts come up every year, on the other hand the thesaurus cannot supply sufficient new terms because of its delayed revision.

5.2 Recommendations for improving indexing consistency

In this study the reasons for various categories of inconsistencies have been specifically examined. Some categories of the inconsistencies have few remedies. The data of the study indicate that indexers have more difficulties in assigning topical terms than the other categories of terms. This category of inconsistencies is attributed to many factors. It may be due to indexers' different knowledge backgrounds and their different experiences with subject indexing, and therefore they have different interpretations of the subject content or focus of documents. Sometimes, just because of following different policies or thinking from different viewpoints, indexers describe the same book in different ways. In this sense, subject indexing is a subjective process. It is inevitable that two indexers interpret the same documents in different ways. However, it does not mean that indexing consistency cannot be improved. Based on the thorough analysis of the reasons for inconsistencies, we can improve it from the following aspects:

Indexers:

During analyzing the reasons for inconsistencies, we notice that a lot of inconsistencies are resulted from indexers' lack of subject knowledge. When indexers are not familiar with the subject areas, they cannot understand the tiny and special differences between two concepts. Thus, they cannot translate their subject analysis into indexing language correctly even if they make a right decision at the stage of subject analysis. Moreover, they are more likely to assign broader terms or classes, when they lack knowledge about the subject field.

However, most indexers are not subject experts. It is not realistic for a library to hire subject experts as indexers, because indexers' familiarity with indexing languages seems more important. Or we should say that these two factors are not independent from each other. If indexers are very familiar with the indexing languages in one subject field, namely knowing well about the index terms or classes in this subject field, then they would acquire the basic knowledge of this subject as well.

Furthermore, indexers should know very well about the indexing rules of the library they work in. When the indexers are not familiar with the indexing languages or the indexing rules, they may easily make mistakes. For instance, they may pay attention only to the name of a class and fail to notice its context in the overall hierarchy, or they cannot correctly conduct a number-building process, or they may choose a broader term or class rather than a co-extensive one, etc.

Obviously, it is very necessary for a library to spend more time and money to train indexers. Making them more familiar with the subject fields, the indexing languages and the indexing rules will to some extent benefit indexing consistency.

Indexing policies:

From the analysis, we can notice that the lack of unified or detailed indexing policy brings about inconsistencies as well. Different library systems may adopt different indexing policies due to their own intended reader groups or the characteristics of their collections. However, in most cases, they have the same general policies due to the nature of subject indexing or because there are international and national subject indexing standards. Thereby, in general, they have many policies in common. In fact, it is not wise to require all libraries to have the same detailed indexing policies. That is to say, the inconsistencies resulted from applying different indexing policies can not be remedied. However, a lot of inconsistencies can be lessened by changing the out-of-date indexing policies or by detailing the unclear ones.

Some indexing policies should be changed because they have been made in the card-catalogue era. With the development of web-based online catalogue they have become unsuitable. For example, more and more interdisciplinary fields are being created and topics in documents are becoming more and more complicated. Describing them requires more than one notation. In the card-catalogue era, the class number is limited to one due to the shelving function of classification systems. However, the situation is changing. In web-based online catalogue, class numbers are distinguished from call numbers, which means that class numbers are separated from shelving function. Actually they are playing a more and more important role in information organizing and searching. The conflict between the old custom (assigning only one notation) and the new requirement (assigning more than one notation) becomes more and more severe over time. In order to resolve this problem, the policy of assigning only one class number should be replaced by a detailed policy of assigning more than one class number.

Moreover, some indexing policies should be detailed. For example, whether or not should indexers assign a broad term after assigning a specific one? According to the data of this study, it happens very frequently that one indexer assigns a broad term after assigning a specific one, while the other indexer does not. The study indicates that we especially need to make a detailed indexing policy for assigning geographic terms. For example, under what kind of situation should a geographic term be assigned? Should a direct entry or an indirect entry be assigned, namely, should the name of the place in question be directly assigned or should the name of a larger geographic place also be assigned? Should the name of a location below the level of city be indexed? If the indirect entry is allowed, how many levels of geographic elements should be contained? If there is a clear and detailed indexing policy of using geographic names and there are common authority control files, it is more likely for indexers to get high indexing consistency in using geographic terms.

In addition, every subject area has its own characteristics and its own problems in indexing. We need more detailed analysis on every subject area, and then we can make different indexing rules for different subject areas according to their own characteristics. By doing this, we can to some extent improve indexing consistency. For example, we need detailed indexing policies for assigning free-text terms in computer science, clear explanations for form terms in education, clear instructions for using general terms in psychology and completed authority files for geographic terms in botany.

Indexing languages:

The bad thesaurus and classification scheme cause low consistency because the indexer can be confused when indexing. To improve indexing consistency, we need to improve the thesaurus and classification scheme first. We cannot leave the thesaurus and classification scheme unrevised for a long time. As mentioned before, the Chinese Thesaurus and the Chinese Library Classification are revised every ten years. The data in this study show that many terms and classes are out of date, which causes a lot of indexing inconsistencies. However, the revision of a thesaurus or a classification scheme is not a simple issue. For example, how often should it be revised? To what extent can it be revised? All these questions should be taken into consideration because the revision can lead to a lot of re-classifying or re-indexing work. Since these questions are beyond the scope of this research, we do not discuss them here. But, some changes in the thesaurus and the classification scheme should be made to improve indexing consistency.

First, we should make an effort to strictly control the vocabulary, especially the synonyms and the terms which overlap a great deal with each other. Regarding the classification scheme, the coordinate classes should be exclusive from each other, and the whole array should be exhaustive of the contents of the class. When it is not possible due to the complex subjects, "see also" reference notes should be added to the scheme. The granularity of terms or classes should be controlled. If the distance among terms or coordinate classes is too small, the difference of terms or the coordinate classes is so subtle that the indexers cannot distinguish them. If it is too large, then the terms or classes are not specific enough to describe specific concepts.

Second, we need to add more notes, especially scope notes and "see also" reference notes. Concerning the thesaurus, some form terms and general terms are not well-defined, e.g. their meanings are vague or their intensions are unclear. This can make indexers confused and consequently cause inconsistencies. Thus, we need to add scope notes to define these terms or explain their meanings. Furthermore, if there are two topical terms in the thesaurus, which are very close to each other, then the tiny and subtle difference should be explained by scope notes so that the indexers with little subject knowledge can distinguish them easily and use them correctly. And a “see also” reference note is also needed because it can help indexers find the appropriate term quickly or help users search documents from different perspectives. But sometimes “see also” references can confuse the indexer as well when the annotations are not clear. Then indexers may interpret the annotations in different ways, which can result in inconsistencies. Thus, detailed annotations are needed for the “see also” references.

The last point is to lessen the indexer’s semantic freedom with term or class choosing. For example, in some cases, there are two alternative classes in the classification scheme, which are used not by force of the classification system but are decided by the library’s indexing policy or by the indexer. This flexibility of using the alternative classes will to some extent cause the inconsistencies. Another example is that almost every number in the schedules can be further extended by notation from the table of Standard Subdivisions without specific instructions. It mainly depends on the indexer’s decision. The freedom of choosing numbers from the table of Standard Subdivisions easily leads to inconsistencies.

5.3 Recommendations for further research in the area

Because of the time limitation some issues have not been investigated or thoroughly discussed in this study. But, they are worthy of further investigation. For instance, further studies are needed to understand the influence of different consistency measures on the results and the practical meaning of each measure. A new measure may be needed. Furthermore, it will be very interesting to know whether consistency can be affected by the language of the document to be indexed. Is it easier to achieve high consistency when indexing Chinese literature?

How to revise the thesaurus or classification scheme based on the frequency of terms or classes would be another interesting area. For example, the terms or classes with high frequency and low consistency ought to be revised or be explained by adding notes. Actually, the document distribution is also useful to revision. Basically, the postings under each term or class should be balanced. When too many postings are gathered under a term or a class, this term or class should be further divided. When too few postings are gathered, it should be combined with other terms or classes.

This study has clearly shown that indexers have big difficulties in assigning form terms and general terms. The same happens concerning the use of the Table of Standard Subdivisions. Should they be further used? If yes, to what extent should they be used and how can they be used appropriately still need to be researched.

References

Angell, Richard S. 1954. Standards for subject headings: A national program. *Journal of Cataloging and Classification* 10 (October): 191-197.

Ankem, Kalyani. 2005. Approaches to meta-analysis: A guide for LIS researchers. *Library & Information science research* 27: 164-176.

Banks, J. 2004. Does the number of subject headings on a bibliographic record affect circulation intensity? *Technical Services Quarterly* 21(3): 17-24.

Bates, M. J. 1977. Factors affecting subject catalog search success. *Journal of the American Society for Information Science* 28: 161-169.

Bates, M. J. 1986. Subject access in online catalogs: A design model. *Journal of the American Society for Information Science* 37(6): 357-376.

Belew, Richard K. 2000. *Finding out about: A cognitive perspective on search engine technology and World Wide Web*. Cambridge [u. a.]: Cambridge University Press.

Bertrand, A., and J. Cellier. 1995. Psychological approach to indexing: Effects of the operator's expertise upon indexing behavior. *Journal of Information Science* 21(6): 459-472.

Booth, A. 1990. How consistent in MEDLINE indexing? A few reservations. *Health Libraries Review* 7(1): 22-26.

Borko, H. 1964. Measuring the reliability of subject classification by men and machines. *American Documentation* 15: 268-273.

Chan, Lois Mai. 1989. Inter-indexer consistency in subject cataloging. *Information Technology and Libraries* 8(4): 349-357.

Chu, C. M., and A. O'Brien. 1993. Subject analysis: The first critical stage in indexing. *Journal of Information Science* 19 (4): 439-454.

CLC Editorial Board. 1999. *Manual on the use of the Chinese Library Classification*. 4th edition. Beijing: Chinese National Library Press.

Cohen, J. 1988. *Statistical power and analysis for the behavioral sciences*. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Collantes, Lourdes Y. 1995. Degree of agreement in naming objects and concepts for information retrieval. *Journal of the American Society for Information Science* 46(2): 116-132.

David, C. et al. 1995. Indexing as problem solving: A cognitive approach to consistency. *Proceedings of the American Society for Information Science* 32: 49-55.

Dominich, Sandor. 2001. *Mathematical foundations of information retrieval*. Dordrecht: Kluwer Academic Publishers.

Drabenstott, Karen Markey, and Diane Vizine-Goetz. 1994. *Using subject headings for online retrieval: Theory, practice, and potential*. San Diego: Academic Press.

Enser, P.G. B. 1995. Pictorial information retrieval. *Journal of Documentation* 51:126-170.

Fattahi, Rahmat. 1998. Library cataloguing and abstracting and indexing services: Reconciliation of principles in the online environment? *Library Review* 47(4): 211–216.

Fried, C. and Prevel, J. J. 1966. *Effects of indexing aids on indexing performance*. Bethesda, MD, General Electric Co., RADC-TR-66-525.

Funk, Mark E., Carolyn Anne Reid, and Leon S. McGoogan. 1983. Indexing consistency in MEDLINE. *Bulletin of Medical Library Associations* 7(2): 176-183.

Gonzalez, Linda . 2005. WebDewey Change Streamlines Workflow. http://www.bcr.org/publications/afl/2005/november/webdewey.html

Gregor, Dorothy, and Carol Mandel. 1991. Cataloging must change! *Library Journal* 1: 42-47.

Harremoës, Peter, and Flemming Topsøe. 2002. Zipf's law, hyperbolic distributions and entropy loss. www.math.ku.dk/~topsoe/ISIT2002Zipf.pdf.

Harris, D. et al. 1966. *The testing of inter-indexer consistency at various indexing depths*. University of Chicago. Graduate Library School. Working paper no. 380-382

Hildreth,Charles R. 1997. The use and understanding of keyword searching in a university online catalog. *Information Technology and Libraries* 16: 6.

Hong, Yi, and Zhang Jin. 1996. The Dewey Decimal Classification in China. *Knowledge Organization* 23(4): 213-215.

Hooper, R. S. 1965. *Indexer consistency tests: Origin, measurements, results and utilization*. IBM Corporation. Bethesda, Md.

Hurwitz, F.I. 1969. A study of indexer consistency. *American Documentation* 20(1): 92-94.

Huton, Michele. 1998. *An assessment of the usefulness of standardized definitions in a thesaurus through interindexer terminological consistency measurements*. Doctoral thesis. University of Toronto.

Iivonen, M. 1990. The impact of the indexing environment on interindexer consistency. *Advances in Knowledge Organization* 1: 259-261.

Jansen, Bernard J., and Udo Pooch. 2001. A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology* 52 (3): 249-250.

Jones, Karen Spark. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* 28: 11-21.

Knapp, P.B. 1944. The subject catalog in the college library. *Library Quarterly* 14: 214-228.

Lancaster, F.W. 1964. Some observations on the performance of EJC role indicators in a medianized retrieval system. *Special Libraries* 55: 696-701.

Lancaster, Frederick Wilfrid. 1968. *Evaluation of the MEDLARS demand search service*. Washington, D.C.: U.S. Department of Health, Education & Welfare.

Lancaster, F.W. 1998. *Indexing and abstracting in theory and practice*. 2nd edition. London: Library Association.

Larson, Ray R. 1991. The decline of subject searching: Long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science and Technology* 42 (3): 210.

Leech, Nancy L., Barrett, K. C., and Morgan, G. A. 2005. *SPSS for intermediate statistics: Use and Interpretation*. 2nd edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Leininger, K. 2000. Interindexer consistency in PsyINFO. *Journal of Librarianship and Information Science* 32(1): 4-8.

Leonard, L. E. 1975. *Inter-indexer consistency and retrieval effectiveness: Measurement of relationships*. Doctoral thesis. Urbana-Champaign, IL: University of Illinois, Graduate School of Library Science.

Leonard, L. E. 1977. *Inter-indexer consistency studies, 1954-1975: A review of the literature and summary of study results*. University of Illinois Graduate School of Library Science, Occasional Papers, no. 131.

Lilley, O. L. 1954. Evaluation of the subject catalog. *American Documentation* 5: 41-60.

Liu, Suqing, and Zhenghua Shen. 2002. The development of cataloging in China. *Cataloging & Classification Quarterly* 35(1-2):137-154.

Losee, Robert. 2004. A performance model of the length and number of subject headings and index phrases. *Knowledge Organization* 31(4): 245-251.

Luhn, Hans Peter. 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development* 2(2): 159-165.

MacMillan, J. T., and I. D. Welt. 1961. Study of indexing procedures in a limited area of the medical sciences. *American Documentation* 12: 27-31.

Mann, T. 1997. "Cataloging Must Change!" and indexer consistency studies: Misreading the evidence at our peril. *Cataloging & Classification Quarterly* 23(3-4): 3-45.

Markey, Karen. 1984. Interindexer consistency tests: A literature review and report of a test of consistency in indexing visual materials. *Library and Information Science Research* 6(2): 155-177.

Marshall, Linnea. 2003. Specific and generic subject headings: Increasing subject access to library materials. *Cataloging & Classification Quarterly* 36(2): 59-223.

Medelyan, Olena, and Ian H. Witten. 2006. Measuring inter-indexer consistency using a thesaurus. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital libraries*. New York: ACM Press.

Middleton, M. R., and Di'Orio, A. 1985. A comparison of indexing consistency and coverage in the AEI, ERIC and APAIS database. *Behavioral and Social Science Librarian* 3(4): 33-43.

Mullison, W.R. et al. 1969. Comparing indexing efficiency, effectiveness, and consistency with or without the use of roles. *Proceedings of the American Society for Information Science* 6: 301-311.

Olson, H. A., and John J. Boll. 2001. *Subject analysis in online catalogs.* 2nd edition. Englewood, Colo.: Libraries Unlimited.

Olson, H. A. and Dietmar Wolfram. 2005. Indexing Consistency and its Implications for Information Architecture. http://www.iasummit.org/2006/files/175_Presentation_Desc.pdf.

O'Neill, E. T., and R. Aluri. 1981a. Evaluation of indexing for large online information systems. *Proceedings of the Annual Meeting American Society of Information Science*. 44th Annual Meeting. 18:104-106.

O'Neill, E. T., and R. Aluri. 1981b. Library of Congress subject heading patterns in OCLC monographic records. *Library Resources & Technical Services* 25: 63-80.

O'Neill, Edward, and Patrick McCain. 1995. Copy cataloguing practices: Use of call numbers by Dewey libraries. *Annual Review of OCLC Research*.

Painter, A. F. 1963. *Analysis of duplication and consistency of subject indexing involved in report handling at the Office of Technical Services, U.S. Department of Commerce*. Unpublished dissertation. Rutgers, The State University of New Jersey.

Pulis, Noelle Van, and Lorne E. Ludy. 1988. Subject searching in an online catalog with authority control. *College and Research Libraries* 49: 526.

Reich, P., and Biever, E. J. 1991. Indexing consistency: The input, output function of thesauri. *College & Research Libraries* 52(4): 336-342.

Rodgers, D.J. 1961. *A study of inter-indexer consistency*. Washington, DC. General Electric Co.

Rolling, L. 1981. Indexing consistency, quality and efficiency. *Information Processing & Management* 17: 69-76.

Saarti, Jarmo. 2002. Consistency of subject indexing of novels by public library professionals and patrons. *Journal of Documentation* 58 (1): 49-65.

Samleske, Roland. 1992. Computer science terminology and the Library of Congress Subject Headings: The case of software. *Technicalities* 12 (2): 11-13.

Sheffler, F.L. 1968. Indexer performance analysis and operations of a document retrieval system. University of Dayton Research Institute, Dayton, OH.

Shoham, S., and R. Kedar. 2001. The subject cataloging of monographs with the use of keywords. *Cataloging & Classification Quarterly* 33 (2): 29-54.

Shubert, Steven B. 1992. Critical views of LCSH—ten years later: A bibliographic essay. *Cataloging & Classification Quarterly* 15(2): 37-91.

Sievert, M. C., and M. J. Andrews. 1991. Indexing consistency in information science abstracts. *Journal of the American Society for Information Science* 42(1):1-6.

Sinnett, J.D. 1964. *An evaluation of links and roles used in information retrieval*. Dayton, Air Force Materials Laboratory, Wright Patterson Air Force Base.

Slamecka, V., and J. J. Jacoby. 1963. *Effect of indexing aids on the reliability of indexers*. Final technical note. Bethesda, MD, Documentation Inc. RADC-TDR-63-116.

Slamecka, V., and J.J. Jacoby. 1965. Consistency of human indexing. In "*The Coming Age of Information Technology*", pp. 32-56. Documentation Inc., Bethesda, MD.

Stevens, M.E. 1965. *Automatic indexing: A state-of-the art report*. National Bureau of Standards, Washington, DC.

Stubbs, Edgardo, Norma E. Mangiaterra, and Ana M. Martínez. 1999. Internal quality audit of indexing: A new application of interindexer consistency. *Cataloging & Classification Quarterly* 28 (4): 53-69.

Subrahmanyam, Bhagirathi. 2006. Library of Congress Classification numbers: Issues of consistency and their implications for union catalogs. *Library Resources & Technical Services* 50 (2): 110-119.

Svenonius, E., and D. McGarry. 1993. Objectivity in evaluating subject heading assignment. *Cataloging & Classification Quarterly* 16(2): 5-40.

Tagliacozzo, R., and M. Kochen. 1970. Information seeking behaviour of catalog users. *Information Storage and Retrieval* 6: 363-381.

Tarr, D., and H. Borko. 1974. Factors influencing inter-indexer consistency. *Proceedings of the Annual Meeting American Society of Information Science*, 37th Annual Meeting, 11: 50-55.

Tell, B.V. 1969. Document representation and indexer consistency. *Proceedings of the American Society of Information Science* 6: 285-292.

Tinker, J.F. 1966. Imprecision in indexing. *American Documentation* 17: 93-102.

Tinker, J.F. 1968. Imprecision in indexing. *American Documentation* 19: 322-330.

Tonta, Yasar. 1991. A study of indexing consistency between Library of Congress and British Library catalogers. *Library Resources & Technical Services* 35 (2): 177-185.

Tibbo, H.R. 1994. Indexing for the humanities. *Journal of the American Society for Information Science* 45: 607-619.

Yu, Holly, and Margo Young. 2004. The impact of web search engines on subject searching in OPAC. *Information Technology and Libraries* 23(4): 168-180.

Zhang, Q.Y., X. Sh. Liu, and D.B.Wang. 1996. Contemporary classification systems and thesauri in China. *Proceedings of the 62nd IFLA General Conference*. http://www.ifla.org/IV/ ifla62/62-qiyz.htm.

Zhang, wenxian. 2004. The development and structure of the Chinese Thesaurus for subject indexing. *The International Information & Library Review* 36: 47–54.

Zipf, G. K. 1949. *Human behavior and the principle of least effor*t. Cambridge: Addison-Wesley.

Zeng, Lei. 1992. Research and development of classification and thesauri in China. In N. Williamson ed.: *Classification Research for Knowledge Representation and Organization*, 409-416. NY: Elsevier, 1992.

Zunde, P., and M.E.Dexter. 1969. Indexing consistency and quality. *American Documentation* 20(3): 259-267.

Lebenslauf von Xu Chen

Angaben zur Person

Name:	Xu Chen
Geburtsdatum:	13.02.1978
Geburtsort:	Luo Yang, Henan Provinz
Staatsangehörigkeit:	China
Geschlecht:	weiblich
Anschrift:	B5, 13 68159 Mannheim
Email:	chenxu@rz.hu-berlin.de

Ausbildung

10. 2003-2007	Promotion der Bibliothekswissenschaft an der Humboldt Universität zu Berlin
09.1999-06.2002	Studium (Master) der Bibliothekswissenschaft an der Peking Universität
09.1995-07.1999	Studium (Bachelor) der Bibliothekswissenschaft an der Peking Universität
09.1983-06.1995	Besuch der Grundschule, Mittelschule und Oberschule in Luo Yang, China.

Berufserfahrungen

06.2002- 09.2003	Angestellte am Beijing Dayang Digital Technology Co. Ltd. in Beijing. Tätigkeit in Informations- und Wissensmanagement
03.2000-10.2001	Nebenjob am Beijing Lanfan Information Technology Co. Ltd. in Beijing. Beschäftigt mit Wissensorganization.

03.1999-05.1999 Praktikum an der Ölfeld Bibliothek in Daqing. Beschäftigt mit Katalogisierung

Projekte

1. Klassifikationssystem (Directory) für eine Suchmaschine
2. Klassifikationsstandard für das ‚National Ministry of Broadcasting and Television'
3. Metadata Standard für das ‚National Ministry of Broadcasting and Television'

Veröffentlichungen

Xu Chen. 2007. The Relationship between Indexing Consistency and Specificity. *Proceedings of International Conference on Future of Knowledge Organization in the Networked Environment.*

Xu Chen. 2002. Using facet classification to organize the World Wide Web. *Library and Information Service*.

Sprachkenntnisse

Deutsch: gut in Wort und Schrift (956 Unterrichtstunden am Goethe Institute Peking, Grundstufe III, Mittelstufe II/III)

English: sehr gut in Wort und Schrift (National College English Grade-6)

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass die vorliegende Dissertation von mir selbst und ohne unzulässige Hilfe Dritter verfasst wurde, auch in Teilen keine Kopie anderer darstellt und die benutzten Hilfsmittel sowie die Literatur vollständig angegeben sind.

Mannheim, August 2007

Xu Chen