

# **An Introduction to subject indexing**

*Second edition*

**A G BROWN**

*in collaboration with*

**D W LANGRIDGE AND J MILLS**

CLIVE BINGLEY



LONDON

# **An Introduction to subject indexing**

*Second edition*

**A G BROWN**

*in collaboration with*

**D W LANGRIDGE AND J MILLS**

CLIVE BINGLEY



LONDON

First published in 1976 in two volumes  
This second edition first published 1982  
by Clive Bingley Ltd, 16 Pembridge Road, London W11 3HL  
Set in 10 on 11 point Press Roman by Allset  
Printed and bound in the UK by  
Redwood Burn Ltd, Trowbridge  
Copyright © A G Brown 1982  
All rights reserved  
ISBN: 0-85157-331-2

**British Library Cataloguing in Publication Data**

Brown, A G

Introduction to subject indexing. — 2nd ed.

1. Subject cataloguing — Programmed instruction

2. Classification — Programmed instruction

I. Title II. Langridge, D W III. Mills, J

025.4'7 Z695

ISBN 0-85157-331-2

## PREFACE TO THE SECOND EDITION

This book was first published in two volumes in 1976. It was developed from work initially carried out under a research project, funded by the then Office for Scientific and Technical Information, designed to investigate the applicability of programmed instruction techniques in the teaching of practical subject indexing. This project was conducted at the School of Librarianship, the Polytechnic of North London with cooperation from the College of Librarianship Wales.

Many programmed texts in the field of subject indexing concentrate upon the translation stage of indexing aiming to impart skills in the use of particular indexing languages. As we stated in the first edition, our intention is to offer a course of instruction which attempts to present an integrated view of some of the basic principles of subject indexing, not to impart a high degree of familiarity with any particular system.

Thus we do not concern ourselves with the Colon Classification and UDC as ends in themselves but to demonstrate the use of classification schemes and the employment of the principles of subject analysis dealt with earlier in the course. In a similar vein we do not treat chain procedure in pre-coordinate index construction primarily as a practical skill to be mastered for its own sake. We believe that an understanding of this technique provides a sound basis for a wider appreciation of pre-coordinate indexing practices and of different types of pre-coordinate indexes.

In this second edition the existing text has been re-structured into one volume. To this we have added a section dealing with the fundamental characteristics of post-coordinate indexing which builds upon the concepts covered earlier.

Rapid developments have taken place over recent years in the field of subject indexing, notably in the sphere of computer applications. The situation now facing the entrant into the areas of library and information work is one of diversity and change. These characteristics are likely to intensify. We feel that it is now probably more important than ever that the beginner is made aware of basic principles which underly this diversity. It is with such principles that this course attempts to deal in the hope that the reader will be better prepared to encounter indexing and retrieval practices lying outside its particular scope.

A G BROWN

School of Library, Archive and Information Studies  
University College London

July 1981

## CONTENTS

---

### Preface

### *Frame*

<i>Section 1</i>	Introduction to subject indexing	1
<i>Section 2</i>	Subject analysis	37
<i>Section 3</i>	Elements of practical classification	131
<i>Section 4</i>	Universal Decimal Classification	203
<i>Section 5</i>	Classified catalogue	270
<i>Section 6</i>	Alphabetical subject catalogue	322
<i>Section 7</i>	Post-coordinate indexing	353

### Appendix

### Index

## SECTION 1: INTRODUCTION TO SUBJECT INDEXING

The principal function of any library is to make the information it contains available to the library users at their request. In order to fulfil this function, the information which is stored in the library must be recovered, or retrieved, from the store. The process of recovering or retrieving information is called, quite simply, *INFORMATION RETRIEVAL*.

*INDEXING* is once of the activities—the major one, in fact,—which supports the process of information retrieval.

In this course we are concerned with that very important part of indexing called *SUBJECT INDEXING*.

In particular, the aim of the course is to teach some of the *practical* skills required in subject indexing, ie some of the techniques of *PRACTICAL SUBJECT INDEXING*.

In order that you should be able to perform these required skills with greater understanding, and therefore with greater competence, selected elements of the theory of subject indexing will be included.

This course is not presented in the form of a conventional textbook, and in most cases you will not read the pages consecutively. Most pages are divided into individually numbered frames. On many pages you will be asked a question. The answer you give or select will direct you to the next frame you should read. In this way, the material is adjusted to your own particular requirements.

For example, having read the above paragraphs, what is the aim of this course?

To teach the complete theory and practice of  
subject indexing

— frame 6

To teach some of the theory and practice of  
subject indexing

— frame 4

2 (12)

No, this is an incorrect answer. If the library catalogue contained only *one* substitute for each document and arranged these in the *same* sequence as the shelf order for documents, it would simply duplicate shelf order. This duplicate order would have the same inherent limitations as the shelf arrangement of documents.

Return to frame 12. When you have re-read the frame, select the correct answer and proceed with the course

### 3 (5)

Your answer: a library fulfils its function of information retrieval by maintaining some system for the recovery of documents from its collection. Quite right.

No matter how *large* the collection, the library is of little value if it is unable to retrieve the right documents as and when they are required. To do this it must maintain an *information retrieval system*.

When documents relevant to a request have been located, a *MATCH* has been achieved between the information requested and the information retrieved.

In other words, the *information supplied* in the document, or documents, *MATCHES*, to an acceptable degree, the *information demanded* by the user.

To achieve a successful match is the central objective of information retrieval.

In relation to a given request for information, under what conditions can this objective be considered successfully achieved?

If relevant documents are contained in the collection — frame 22

If relevant documents are located in the collection — frame 9

---

### 4 (1, 6)

(NB: Numbers given in brackets refer to frames which have led to the one you are reading. This is to help you retrace your steps if necessary.)

You are correct in your answer. The aim of the course is to teach you only *some* of the theory and practices of subject indexing.

In particular, our concern is with the problems of the subject indexing of a given document in a given library system. We shall begin by defining two frequently recurring terms—library and document.

There are many kinds of library and information centre each fulfilling their individual purposes. The public library, the university library, the library of a commercial firm, for example, each serve the various needs of differing groups of users.

The fundamental characteristic common to *all* libraries is that they are stores of information.

This information is contained in information carriers, which vary widely in their physical forms, eg books, films or gramophone records.

In this course we are concerned not with any particular kind of library nor with any particular form of information carrier. We shall thus use the terms

*DOCUMENT* to mean any form of information carrier  
and

*LIBRARY* to mean any collection of documents  
Continue on the next frame.

## 5 (4)

A library acquires documents because they contain information of the kind that is likely to be of interest to its users. At this stage we need not worry about the precise nature of this information.

Information retrieval is the process of satisfying the requests of library users by providing them with relevant information contained within the library. It is the principal function of a library.

The term information retrieval usually implies *document retrieval*. That is, the satisfaction of a request for information by retrieving a document, or documents, which will contain information relevant to that request.

As such it is usually distinguished from *DATA RETRIEVAL*—the satisfaction of a request for information by providing the information as a direct answer to the question.

We shall use the following definition of *INFORMATION RETRIEVAL*: *The recovery of documents from a given collection which are relevant to a request.*

By what means does a library fulfil its function of information retrieval?

By acquiring a large collection of documents — frame 13

By maintaining some system for the recovery  
of documents from its collection — frame 3

---

## 6 (1)

(NB: Numbers given in brackets refer to frames which have led to the one you are reading. This is to help you retrace your steps if necessary.) No, this is not quite correct. To attempt to teach you the *complete* theory and practice of subject indexing is beyond the scope of this course.

We shall concentrate on only some of the skills required in practical subject indexing.

Moreover, it was stated that only selected elements of subject indexing will be included. Those elements which will contribute directly to the execution of these skills with greater competence.

Proceed to frame 4 and continue with the course. Remember to read each frame *very carefully*.



7 (10)

Your answer: only set A constitutes a class of documents. No, this is not correct.

You are right in thinking that set A constitutes a class. All the documents in this class share at least one characteristic in common—they are all *about* the French Revolution. Their common characteristic is their *subject content*.

However, there is also a common characteristic possessed by the documents in the other two sets. Your answer does not go far enough.

Read frame 10 once more and reconsider the question.

---

8 (10)

You say that only sets A and B constitute classes of documents. This is not true.

You have correctly observed that all the documents in set A are about the same subject, and all the documents in set B are written by the same author. These are therefore two classes of documents defined respectively by the common characteristics of subject content and authorship.

But what about set C? All the documents in this set are published by the same publisher, the Cambridge University Press. This is also a class of documents for each document possesses at least one characteristic in common—its publisher.

We said that, in information retrieval, a class is a set of documents which share *some* property or characteristic in common.

Your answer should have been: *all* the sets constitute classes.

Go to frame 14, and continue with the course.

9 (3)

Your answer: If relevant documents are located in the collection. Yes, you are correct in saying that only under this condition has a match been successfully achieved.

The possession of relevant documents, does not, itself, imply a match in terms of information retrieval. To achieve a match we must be able to locate these documents within the collection.

In order to locate documents relevant to a request, the collection, that is the information store, must be examined or *SEARCHED*.

To illustrate the process of searching let us take a simple example of practical information retrieval. Suppose you wish to borrow a book from your public library about '*Programmed instruction*'.

Now the collection of documents in the public library covers the whole range of knowledge. In the attempt to satisfy your particular request for information, you would not expect to have to examine every document in that collection. You would not expect to search the entire information store.

You would ignore the documents about history, engineering, bio-chemistry etc. You would confine your search to those documents about '*Programmed instruction*'. Within this group of documents, probably shelved with other documents about education, you would most likely find one or more relevant to your needs. At this point you would have achieved a *match* between information demanded and information supplied.

It is obviously impracticable to search the entire information store in the satisfaction of a particular request for information.

*The basic principle of information retrieval is to search only a limited part of the store in response to each request, that part which is potentially relevant to the request.*

In the above instance, your request was for information about a particular subject, '*Programmed instruction*'. You would therefore limit your search to that part of the store potentially relevant to this subject.

A limited part of the store, such as this, is an example of a *CLASS*. Continue on the next frame.

10 (9)

**A CLASS IS A SET OF THINGS WHICH SHARE SOME PROPERTY, OR CHARACTERISTIC, IN COMMON.**

For example, violins, cellos, harps and guitars are all musical instruments which produce sound through the bowing or plucking of strings. This particular medium of sound production is a characteristic which they all possess, it is a *common characteristic*.

We can thus say that violins, cellos, harps and guitars form a part of that *class* of musical instruments called *stringed instruments*.

Other classes of musical instruments defined by the characteristic of sound production would be percussion instruments, wind instruments, etc.

In information retrieval, a class is a *set of documents* which share some property or characteristic in common.

For example, the set of documents about '*Programmed instruction*' forms a class by virtue of sharing the common characteristic of *subject content*. Each document is about '*Programmed instruction*'.

Look at the following set of documents.

A A set of documents about the French Revolution

B A set of documents written by Thomas Carlyle

C A set of documents published by the Cambridge University Press

Which of the following statements is true of these sets

Only A constitutes a class — frame 7

Only A and B constitute a class — frame 8

All constitute classes — frame 14

11 (16)

You say the document will not be automatically retrieved. Quite right.

If '*The aims of education*' is classified by its subject content and arranged with other documents about education, it cannot at the same time be displayed as belonging to that class of documents written by A N Whitehead.

This illustrates the major weakness inherent in the shelf arrangement of documents as an aid to their retrieval.

Although a document can be usefully regarded as belonging to a number of classes, it can only be located in *one* of these.

The only way in which a library could simultaneously display a document in all its potentially relevant classes, would be to buy several copies of the document. A policy of multiple acquisition of all documents is obviously uneconomic.

There are other weaknesses in shelf arrangement for retrieval purposes:

In any given class, not all the documents belonging to that class are necessarily present at any one time, eg some on loan from the library.

In any given class, not all the documents belonging to that class can necessarily be stored in any one place in the library, eg large size books, pamphlets etc often require special storage facilities.

Continue on the next frame.

## 12 (11)

The physical arrangement, or 'shelf arrangement', of documents is of limited value as a tool for information retrieval. To help overcome these limitations *records* are made of the documents held in the library.

Such records come in a variety of physical forms. One commonly used form is the 5 × 3 inch catalogue card upon which the required data is typed.

Each record of a document contains a description of the document—its author, title, place of publication, publishers, date of publication etc. The record can thus act as a *SUBSTITUTE* for the document.

These document substitutes are arranged together to form a *LIBRARY CATALOGUE*. The catalogue forms a complete record of all the documents held in the library.

For each document several substitutes can be made and these can be arranged in several different sequences, or ways, in the catalogue.

The principal limitation in the shelf arrangement of documents as an aid to information retrieval is that a document can only be located in any one class at any one time.

How can the library catalogue overcome this limitation?

By containing one substitute for each document  
and arranging these in the same sequence as the  
shelf order of documents

— frame 2

By containing more than one substitute for each  
document and arranging these in sequences  
differing from shelf order

— frame 20

---

## 13 (5)

You think that a library will fulfil its function of information retrieval by acquiring a large collection of documents.

A large collection of documents may well contain a great deal of useful *information of value to the library users*. However, a large collection of documents is of little value in itself unless documents can be recovered when needed. Information retrieval is the process of recovering documents when they are required.

Read frame 5 again. Then reconsider the question and continue with the course.

14 (7, 8, 10)

You say that all the sets of documents named constitute classes of documents. You are correct.

All the documents in each set share at least *one* characteristic in common. They are all either about the same subject, written by the same author or published by the same publisher. Each of the three sets is thus a class of documents, each class being defined by a different characteristic.

*CLASSIFICATION* is the activity of forming classes. In the particular context of information retrieval, this activity is sometimes referred to as *LIBRARY* or *BIBLIOGRAPHIC CLASSIFICATION*.

At its widest interpretation, the classification of documents can be taken to mean the definition of classes by any of a number of possible characteristics. In this broad sense it is the basic activity in all aspects of information retrieval.

Continue on the next frame

## 15 (14)

The whole process of information retrieval is initiated by a request for information.

Now these requests are couched in a variety of ways which express differing approaches to information needs.

For example, there are requests for information on *named subjects*. 'What has the library got *on programmed instruction*', 'Have you any books *about classification*?' etc.

There are requests for a document, or documents by a *named author* or from a *named publisher*.

These kinds of requests, and many others, all exhibit valid approaches to the expression of information needs by library users.

You have seen that the basic principle in information retrieval is to search only a limited part of the store in response to each request. That part of the store which is potentially relevant to that request.

In other words, we search that *class of documents* which is potentially relevant to the request.

It is obviously useful to have the documents themselves arranged into classes. The searcher can then go to the library shelves and examine the relevant class.

If you wish to retrieve a document about a named subject it would be useful to have the documents arranged in classes defined by their subject content. We presumed this principle of organization in the case of searching the public library for a document about programmed instruction.

If you wanted to retrieve a document by a named author, it would be useful to have the documents arranged in a classified order defined by their authorship.

In fact, it would be useful to have the documents arranged into classes defined by all the characteristics by which they are sought.  
Continue on the next frame

---

## 16 (15)

So far, so good. Now take the case of an individual document, say '*The aims of education*' by A N Whitehead.

This document is relevant to a request for information about education. It is also relevant to a request for documents written by A N Whitehead.

Suppose the library possesses only one copy of this document and it has been arranged on the shelves along with other documents about education. It has been arranged in a class defined by its subject content.

A request is placed for documents by A N Whitehead. What will be the result if the library relies on shelf arrangement as the sole aid to its information retrieval?

The document will be automatically retrieved — frame 24

The document will not be automatically retrieved — frame 11

17 (30)

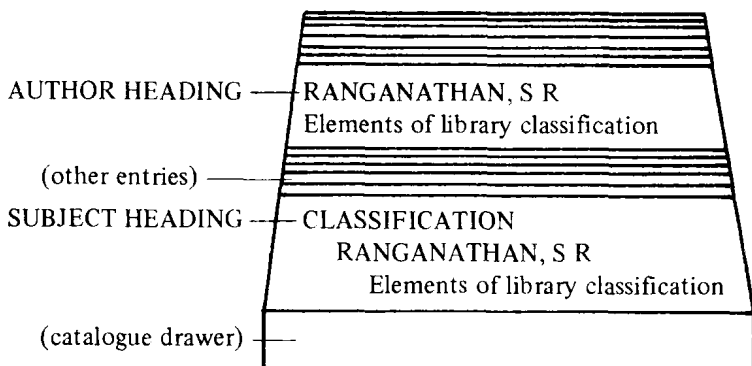
You say that the heading on each document entry will consist of a description of a *class* to which the document belongs, rather than a full description of the *document*. Yes, you are right.

Take, for example, the document '*Elements of library classification*' by S R Ranganathan.

Let us say that this document can usefully be regarded as belonging to two classes—one defined by its *authorship*, one by its *subject content*.

It will thus receive two entries in the catalogue—one having an *AUTHOR HEADING*, one a *SUBJECT HEADING*.

These entries could be arranged in a card catalogue in a single alphabetical sequence as illustrated below.



On searching the catalogue for information on *classification*, this document will be indicated as relevant.

On searching for documents written by *S R Ranganathan*, this document will also be indicated as relevant.

Because the catalogue can contain more than one substitute for each document it is said to allow for *MULTIPLE ACCESS* to documents—ie access via all the different characteristics by which a document is liable to be sought and which define its class membership.

Later in this course we shall have much more to say about library catalogues, their structure and function.

Continue on the next frame



18 (17)

For the present it is sufficient to note that, as aids to the retrieval of documents, they have the following advantages over shelf arrangement:

1 They allow for the multiple access to documents.

2 In any given class, all the documents contained within that class can be represented at one place in the catalogue, irrespective of their shelf location in the library.

3 At any given time, all the documents in a class will be represented in the catalogue as being contained within that class, even if they may be temporarily removed from the library eg on loan.

Continue on the next frame

---

19 (18)

By means of the arrangement of document substitutes in library catalogues, and also by the arrangement of documents themselves, it is possible to point out, or indicate, classes of documents. In doing this we facilitate the process of information retrieval.

In the context of information retrieval, an *INDEX* is some kind of physical mechanism, or tool, which serves to indicate to the searcher those parts of an information store which are potentially relevant to a request.

A library catalogue is a form of index.

Would it be true to say that the shelf arrangement of documents is also a kind of index according to our definition of the term?

Yes — frame 25

No — frame 21

---

20 (12)

Your answer: by containing more than one substitute for each document and arranging these in sequences differing from shelf order.

Correct. It is by this means that the library catalogue can overcome the limitation of one document, one location which is inherent in shelf arrangement.

Document substitutes eg catalogue cards, are relatively cheap to produce. It is therefore quite feasible to have several such substitutes for each document. Given this condition, each document can be displayed simultaneously in a number of classes via the medium of these substitutes arranged in the catalogue.

Each document substitute is said to constitute an *ENTRY* for that document in the catalogue.

Each document entry serves to indicate the inclusion of that document within a class of documents and so facilitates its retrieval.