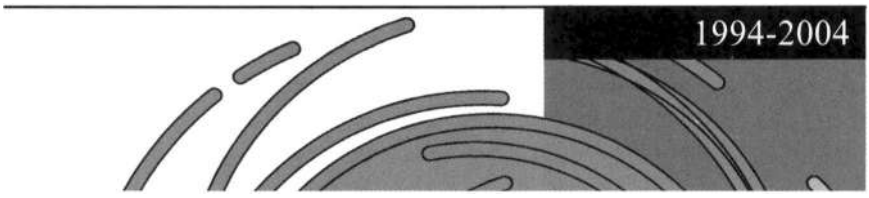
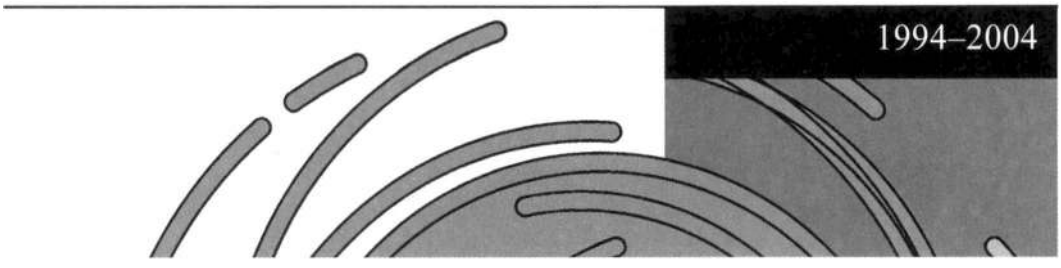
7526

INDEXING FOR SOUTHERN AFRICA



INDEXING FOR SOUTHERN AFRICA

A Manual Compiled in Celebration of ASAIB's First Decade



Editors
Jacqueline A Kalley
Elna Schoeman
Marlene Burger

UNIVERSITY OF SOUTH AFRICA, PRETORIA

© 2005 University of South Africa
First edition, first impression

ISBN 1-86888-380-9

Published by Unisa Press
University of South Africa
P O Box 392, 0003 UNISA

Cover and layout design: Elsabe Viljoen
Electronic origination: Compleat
Printed by Interpak, Pietermaritzburg

© All rights reserved. No part of this publication may be reproduced in any form or by any means – mechanical or electronic, including recordings or tape recording and photocopying – without the prior permission of the publisher excluding fair quotations for purposes of research or review.

CONTENTS

Introduction ix
JA Kalley

Part One: THEORETICAL ISSUES

1 History of Indexing in South Africa: A Brief Overview 3
R Musiker

2 Indexing in Context 7
JC Theron & GR Davis

3 Conceptualisation: What is it all About? 20
I Fourie

4 Characteristics and Evaluation of a Good Index 29
A-M Arnold

5 Organisation Policy 44
M Burger & I Fourie

Part Two: GENERAL APPLICATIONS

6 Verbal Subject Description 53
I Fourie & M Burger

7 Book Indexing 68
I Fourie & M Burger

8 Embedded Indexing for Creating Book Indexes 86
E Lastovica

9 Periodicals Indexing 92
A. Kotze

Addendum: Principles of Periodical Indexing 103
M Burger

10 Automatic Indexing 111
P Underwood

11 Web Indexing 117
M du Preez
12 Metadata and Indexing 129
A Watkins
13 Abstracting 141
M Burger
14 Thesaurus Construction 159
M Burger
15 Technical Presentation of Book Indexes 179
U Reddy & O Jagarnath
16 Standardisation 189
P Duncan

Part Three: SPECIAL FIELDS

17 Africana 197
Y Garson, P Duncan & JC Theron
18 Antique Maps 210
Y Garson
19 Archives 216
M Pickover, C Archibald & K Abbott
20 Education 221
S le Roux
21 Environmental Issues 226
M Pretorius
22 Health Sciences 233
E Robertson
23 Labour 238
J Shier
24 Languages 243
SM Willet
25 Law 250
NM Ferreira

26 Literature for Children 257
TB van der Walt & K. McGuirk
27 Museum Objects 268
ME Schulze
28 National Bibliographies 285
D Ambrose
29 Newspapers 292
C Merrett
30 Politics and International Relations 297
JA Kalley & E Schoeman

Part Four: TRAINING

31 Training Opportunities 307
M Burger

List of Contributors 323
Index 326

INTRODUCTION

Jacqueline A Kalley

Indexing is a strange occupation and one that requires a special type of person – one who thrives on creating order but also one who is blessed with an essential mix of an informed background, a sparkling curiosity, the ability to make lateral connections and the patience to untangle the strands of the narrative and then weave them anew into a finding instrument that is logical, succinct and objective. To create this retrieval tool, the indexer has to plunge into another's way of thinking and often with very little time at hand, familiarise oneself with the intricacies of a subject that is often new and complex. How is this achieved? Having practical experience is often cited as essential for a good indexer, but how does one acquire this experience? At the expense of the author? By producing a series of dubious offerings? Or by seeking the assistance of a manual, carefully crafted to assist the user in overcoming the twin vagaries of ignorance and inexperience by consulting a practical guide produced by experts in the field?

ASAIB has attempted to fill this gap in the Southern African environment by the publication of this book. It is the highlight of ten years of service to the indexing community in providing workshops, conferences and their published proceedings, bi-annual newsletters, training and the collegial support necessary to the relatively small body of indexers and bibliographers in the Southern African region. This manual reflects a variety of levels and approaches and could benefit students of information science, occasional indexers, anyone interested in becoming an indexer, and information professionals in need of a review of indexing techniques. For the practising indexer, this is a source of reference, as it brings together areas of knowledge that are the foundations for creating a good index.

ASAIB began, as have many good things in the information science field, with an idea based on the work done by the Society of Indexers in Great Britain. Almost to the day, ten years ago on 21 September 1994 a meeting was held by interested people at Esselen Park to establish ASAIB. Office bearers were elected, affiliations discussed and the decision was taken to establish an ASAIB newsletter. Reuben Musiker became ASAIB's first chairman of the Executive Committee (Exco). In accordance with ASAIB's Constitution he served for two terms, followed by Jacqueline Kalley and thereafter Marlene Burger. From a small beginning ASAIB has burgeoned to boast a membership of some 140 members, branches in Cape Town and KwaZulu-Natal, its own web page (www.asaib.org.za), and a directory of practising indexers. It has contributed towards the information science profession by the holding of an annual conference followed by the publication of the proceedings, numerous training workshops and a bi-annual newsletter. In the compilation of this manual, it aims to provide further service to Southern Africa's indexing community.

The manual is divided into four parts. The first part is devoted mainly to theoretical issues and guidelines. Set against the background history of indexing in Southern Africa, the work then attempts to contextualise indexing as an indispensable tool in the retrieval of universal knowledge, and, more specifically, in the field of information organisation and retrieval. A consideration of the complexities of conceptualisation is offered and different

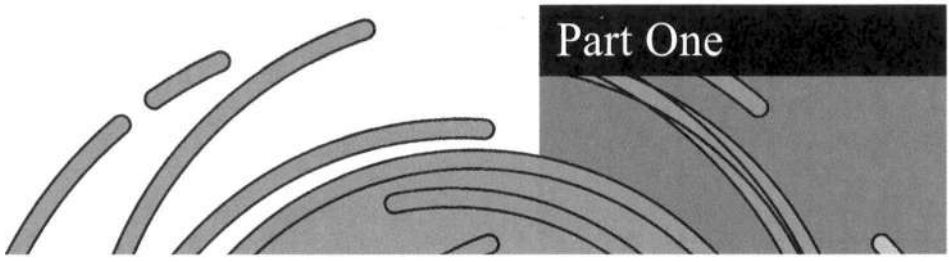
views on the subject are presented. This is complemented by the characteristics and an evaluation of a good index; the former is a vital step towards mastering the quality and competence in the production of indexes. The importance of having an indexing policy is also considered.

Verbal subject description is analysed in the second part, leading to a discussion on the intricacies of book indexing. The reader is familiarised with the technical presentation and editing procedures of an index. Knowledge of periodical indexing is also a vital component of the discipline. The *Index to South African Periodicals* is used as an example of a periodicals subject indexing database.

Developments in utilising electronic techniques for indexing then become the focus of analysis. This commences with embedded indexing, a paperless process that has been defined as the inserting, or embedding, of index entries into an electronic document with the software that was used to create the document. Unlike automatic indexing, a detailed description of which subsequently follows, the process is done manually by an indexer. The meaning of web indexing (or online indexing) is explained and some web indexing decisions and questions are considered. A brief explanation of HTML is given in order to introduce meta-tags as a means to control the way in which search engines index a website. The concept of metadata is further expanded upon together with an overview of the various types of metadata. The importance of standards is emphasised, with details of the Dublin Core provided as an example of a widely used metadata standard. Steps in the creation of metadata in a digitisation project conclude the chapter. As a means of making the mass of information more manageable, and thus providing a short, accurate representation of the intellectual content of the book, general rules and guidelines in creating an abstract are presented. Part two is appropriately rounded off by a discussion on the standards that apply to indexing, especially ISO 999 1 900 on their content and presentation.

Part three of the work concentrates on special fields in order to assist the indexer with practical information on the indexing of a specific subject. These are wide ranging in scope and every attempt has been made to cover a variety of subjects. Information on the indexing of scientific information proved to be difficult to obtain and should perhaps form a companion volume to this work. Indexing of the Health Sciences goes some way towards rectifying this lacuna. The special fields are arranged in alphabetical order, all of which are compiled by well-known practitioners. These include Africana, antique maps, archives, education, environmental issues, health sciences, labour, languages, law, literature for children, museum objects, national bibliographies, newspapers, politics and international relations, illustrating the diversity of material covered.

The manual concludes with one of the most vital areas in which ASAIB has attempted to assist the Southern African information community and that is in training. Indexing training opportunities are assessed at university, technikon and indexing organisations. Indexers need a good general education, advanced subject knowledge and specialised acumen on indexes and indexing. This manual, with its peer-reviewed and informative contributions will make a profound impact by assisting the indexers of Southern Africa and beyond to hone and perfect their skills and make a contribution to the intellectual life of the region.



THEORETICAL ISSUES

1 HISTORY OF INDEXING IN SOUTH AFRICA: A BRIEF OVERVIEW

R Musiker

Abstract

In contrast to bibliography, indexing in Southern Africa does not have a formal history. Most indexing developments are of recent origin. Principal ventures mentioned in this overview relate to periodicals and newspaper indexing, the outstanding example being the *Index to South African Periodicals*, which has been in existence for over 60 years. Attention is paid to indexing as a profession, in which connection the Association of Southern African Indexers and Bibliographers (ASAIB) has played a leading role. Examples are provided of some of the major indexes in a wide range of subject fields. The overview concludes with a list of important electronic indexes that have appeared in recent years.

Introduction

Indexing in Southern Africa is as old as the earliest books themselves, although not, of course, as a formal discipline. Indexing as an art in the subcontinent only came into its own in the second half of the twentieth century. This brief historical review of the principal highlights in this field is considered first in terms of periodicals and newspapers, and thereafter from the perspective of monographs and other categories of publications.

Bibliography and indexing have always been, and still are, closely related disciplines. The history of bibliography in South Africa has been comprehensively documented in R Musiker's monographs (Musiker 1996; 1997). In stark contrast, the history of indexing in South Africa has not been covered in similar fashion because indexing has not been considered as a *formal* discipline until recent times. Indexing has been a neglected field, in fact the 'Cinderella' of the two disciplines.

As far as retrospective South African bibliography is concerned, Sidney Mendelssohn's *South African bibliography* (1910) is a landmark work, and in fact it is still considered to be one of the highlights of retrospective bibliography in South Africa. Regrettably, the two-volume work, despite having many admirable features, was devoid of an index.

The successor to Mendelssohn was *A South African bibliography to the year 1925,* and this is the definitive standard reference work for tracing publications on, or published in South Africa prior to 1925. The publication of this six-volume work between 1979 and 1998 culminated in the publication of an index volume (Titles and Subjects) compiled by Fransie Rossouw in 1998. The index was a meticulous compilation of the highest standard and excellent in every respect. The compiler was awarded the SAILIS Award for Indexing.

The serious gap in the South African bibliographic record between the commencement of the *South African national bibliography* in 1959 and the terminal date of *A South*

African bibliography to the year 1925 was filled by the publication of the *Retrospective South African national bibliography 1926-1958*. This work contains 35 000 entries and is equipped with an admirable index.

Indexes to Periodicals

The importance of a periodicals index in making a nation's writing fully accessible to research workers and librarians needs no restatement here. Periodicals without indexes are rendered virtually useless. Considerable attention has therefore been paid by the library profession in South Africa to the matter of an adequate index to periodicals.

In 1966 the Johannesburg Public Library launched a project to index the periodicals literature of South Africa for the years 1900-1939. The 60 periodicals that were analysed represent a considerable body of knowledge. The National Library of South Africa has published this cumulation on microfiche.

The *Index to South African Periodicals,* which covers the period from 1940 to date, is an example of a most successful enterprise in this field. The Johannesburg Public Library was responsible for the work for the first four decades, excepting the first two years.

In the 1980s, the State Library assumed responsibility for the Index. The State Library in Pretoria and the South African Library in Cape Town merged to form the National Library of South Africa in November 1999. The Index is produced in collaboration with other information organisations that are contracted to the National Library to do the indexing. In terms of this arrangement, the Council for Scientific and Industrial Research, the University of South Africa Law Library and the Institute for Theology and Religion each indexes periodicals for a broad subject area.

The index covers 800 of the more important current and discontinued South African periodicals. The database for this Index contains more than 440 000 entries and is updated weekly by 2 800 entries each month. Scholarly journals are fully indexed and the more general and popular ones are selectively indexed. Certain reviews and original literary contributions are also included. Decennial volumes in hard copy exist for 1940-1949, 1950-1959, 1960-1969 and 1970-1979, but these are out of print. They are available for consultation at the National Library. A cumulative microfiche edition for 1940-1986 has been published.

The *Index to South African Periodicals* (ISAP) has been accessible on SABINET since 1987 (updated weekly), as well as the National Inquiry Services Centre (NISC). The latter publishes anthology databases containing ISAP records on CD-ROM, quarterly, as well as their BiblioLine website (reflecting weekly updates). The National Library makes the ISAP database available on the Internet via ISAPOnline.

Indexes to Newspapers

Retrospective newspaper indexing is less than satisfactory, as there have only been one or two isolated efforts aimed at remedying the situation.

The Index for 1831-1921 to the *Grahamstown Journal*, compiled at the Cory Library for Historical Research, Rhodes University, Grahamstown, deserves special mention.

1831-1876 is a personal names index only, whereas 1876-1921 was broader in scope and included biographical items in the newspaper's news matter.

The Cape Town English press index compiled by Peter Coates has so far covered the years 1871-1876, with 1877 scheduled for the near future. The compiler plans to complete the decade 1871-1881 in due course because of the historical significance of these years. The indexes for 1871-1877 total approximately 8 000 pages and are available for sale on microfiche from the National Library of South Africa (NLSA) and for consultation on hard copy at NLSA, Cape Town Campus.

South African Jewish History

The pioneering bibliographic and indexing work of S. A. Rochlin, first archivist of the South African Jewish Board of Deputies resulted in two landmark products, both of cardinal importance in South African Jewish history. In the first instance, Rochlin compiled a set of abstracts to the South African content of the *London Jewish Chronicle* covering the years 1859-1910. The *London Jewish Chronicle* is the premier, and at times the only, source of important information about South African Jewry.

To complement this tool, and to facilitate information retrieval from it, he then compiled a subject index to the abstracts. This index will be published in the near future. Rochlin also compiled abstracts and indexes to the Jewish content of early South African newspapers, for example *Standard* and *Diggers' News.*

Indexes to Government Publications

An *Index to South African Government and Provincial Gazettes 1910-1989* was published by the Government Printer. The *Juta-National Library Index to the Government Gazette,* 1990 to date, contains an alphabetical subject index and numerical index.

Book Indexes in Selected Subject Fields

Some outstanding examples of indexes in major South African subjects:

- Religion:

  Laver, MPH. 1972. *Index to the South African edition of the Anglican book of common prayer.* Johannesburg: University of the Witwatersrand, Department of Bibliography, Librarianship and Typography.

- Race relations:

  Potgieter, PJJS. 1979. *Index to literature on race relations in South Africa 1910-1975.* Boston: Hall.

  South African survey. 1935-. Johannesburg: South African Institute of Race Relations. Formerly titled: *Race relations survey.*

- Geology:
  Bibliography and Subject Index of South African Geology. Pretoria: Geological Survey. Annual.
- Encyclopaedias:
  Standard encyclopaedia of Southern Africa. 1970-1976. Cape Town: Nasou. 12 vols. Volume 12 is Index.
- Yearbooks:
  South African yearbook. 1994-. Pretoria: Government Communications and Information Service. Includes comprehensive index.
- Law:
  Juta's Statutes of South Africa. 7 vols. Volume 7 is Index Volume.
  South African law reports. 1947-.
  South African criminal law reports. 1990-.

The last two titles above are equipped with consolidated indexes.

Electronic Indexes

The following indexes have been published electronically on CD-ROM:

Index to South African periodicals. 1987-.
Index to South African theses and dissertations. 1918-.
Index to Afrikaans literature since 1988-.
Index to South African English literature. 1800-.
South African national bibliography. 1988-.
Waterlit: Databases of 221 000 citations published. 1975-.
FishLit: Index to 50 000 records. 1985-.
South African law reports (1947-) and *South African criminal law reports* (1990-) are both available electronically.

Professionalisation

The establishment of the Association of Southern African Indexers and Bibliographers (ASAIB) in September 1994 to serve the interests of the indexing profession was a very significant event. The Association has organised training courses, workshops, conferences, published numerous monographs and developed a website. It has done all it can to advance the cause and profession of indexing in Southern Africa. An award is made annually for the best bibliography or index. ASAIB is affiliated to the Society of Indexers (United Kingdom).

BIBLIOGRAPHY

Musiker, R. 1996. *South African bibliography.* 3rd ed. London: Cassell. 1st ed 1970; 2nd ed 1978.
Musiker, R. 1997. *Guide to South African reference books.* 6th ed. London: Cassell. 1st ed 1955.

2 INDEXING IN CONTEXT

Christie Theron and Gavin Davis

Abstract

This chapter attempts to contextualise indexing as an indispensable tool in the retrieval of universal knowledge and, more specifically, in the field of information organisation and retrieval. It also attempts to place information science as a discipline among other disciplines dealing with information transfer. It furthermore focuses on the role of indexers and indexing in society as well as on some of the future problems facing indexing.

Introduction

Indexing, indexes and indexers are a part of everyday life in libraries and information centres, and also of our daily lives. It is a natural action to consult an index at the back of the book if looking for information or to see if the book covers a topic of interest. The computer and Internet literate will use a search engine such as Google or Yahoo to look for information. It may be safely said that a lot of people use indexes without even realising that they are doing so. This points to the effectiveness and general acceptance of, or to the 'embeddedness' of indexes in a literate society.

People ask questions like the following: What is an index? What is its purpose? How, if at all, does it fit into scholarly disciplines? This chapter is an attempt to clarify some of these issues. The discussion will sometimes be general, but the purpose is to give an overview of the terrain or activity of indexing. However, the discipline of information science is also briefly discussed in this context. Certain topics are discussed in more detail in other chapters of this book.

Indexing Defined and Described

A few definitions and descriptions of the concept of index and indexing will be useful:

> Index 1. Anything that serves to guide, point out, or otherwise facilitate reference, as: a ... An alphabetical listing of names, places, and subjects included in a printed work that gives for each item the page on which it may be found. .... 2. Anything that reveals or indicates; a sign, a token (*Universal dictionary* 1987:780).

Lancaster (1998:1) describes indexing as an activity 'to construct representations of published items in a form suitable for inclusion in some type of database'. Indexing fits into a broader framework of information retrieval and access to information. It deals, *inter*

alia, with providing access to information in any form, for instance in books, journals, technical documents, websites and other information objects and resources (Bridgeway Career Development 2001).

The *New Encyclopaedia Britannica* (1990:558) defines indexing as 'extracting from [a document] or assigning to it subject and other "descriptors" – words or phrases denoting significant concepts (topics, names) that occur in or characterize the content of a record'. Keenan and Johnston (2000) offer a similar definition when they see indexing as the 'process of analysing the information content of an item and expressing this in the language of a particular indexing system'.

The quoted definitions indicate what an index is or what different purposes it can serve. The first description states that an index is a list of terms, names, places, et cetera, used in the book and showing where these can be found – that is, a specific page number. An index can also refer the user who looks up a term that is not used in the book to the one that is used, by linking the terms in the index. For example:

Puma Smuts *see* Smuts, Isle; or
see also Smuts, Jan Christian

The second important point is that an index acts as a pointer to valid information applicable to the user's query. An analogy would be that smoke points to the existence of a fire, while an entry in a library catalogue indicates where an appropriate book can be found in the library collection.

So an index can be described as an aid in the process of information retrieval, be it from a book, a database or a library collection. The core of the discipline of information science can be described as the activity of bringing relevant information to the user needing the information. In this process, indexing plays a key and essential role. The tools or apparatus designed by the information scientist to achieve this activity form part of indexing and are thus also indexes, the final product of the activity of indexing. However, indexing is also an intellectual process, which requires higher cognitive skills. Index lists in books, bibliographies, library catalogues, indexing and abstracting journals and services, online databases and index lists such as *Index medicus*, search engines on the Internet, guides to archival collections and plant classification systems are all indexes.

Steps in the Indexing Process

The indexer follows two principal steps in the indexing process:

- conceptual analysis
- translation

These appear two very simple steps to follow – but, as most indexers know, they are complex, highly intellectual and often time-consuming tasks. These tasks are often complicated or made even more difficult by outside factors such as time constraints from publishers, financial factors, family circumstances and authors' requirements.

Conceptual Analysis

The conceptual analysis of documents requires analytical skills and subject knowledge, often in a variety of disciplines, to enable the indexer to see multiple viewpoints and to link different concepts within a document (Bridgeway 2001; *Indexing* 2003). See chapter 3 for more information.

Translation

This implies that the concepts identified should be matched to the language of the indexing system used, for instance a classification notation, a search term in a thesaurus or possible search terms that can be used by searchers.

These tasks are based on a theoretical framework, for instance the theory of knowledge creation, its existence, storage, retrieval and use (epistemology, ontology, etc.). These matters will be discussed in the next section. Becoming a good indexer thus requires a sound theoretical schooling. In the present technologically obsessed society, knowledge of technology is important, as well as entrepreneurial, human resources and office management skills. Indexing thus consists of the conceptual analysis of an information record (book, journal article and web page) and the translation of these concepts into the language of an information retrieval system (book index, subject index and library catalogue).

Indexing as a Part of the Broader Field of Information Theory

Indexing as a discipline or an activity does not exist in a vacuum. It forms part of the broader field of information storage and retrieval that in turn is integral to the discipline of Information Science and/or Library and Information Science (LIS). LIS supplies the theories on which the practice of indexing is based. It is, however, a two-way process, as is the case in most theory-practice relationships. Indexing practice also contributes to theory formulation in LIS, and vice versa. However, information science as a discipline is difficult to define and describe. In 1984, Schrader (1984) identified more than 40 different names or terms as well as the same number of definitions for the discipline of information science. Ingwersen (1992) and De Beer (1998) also made valuable contributions to describing the field of information science. For the purpose of this chapter a few definitions will suffice.

Taylor (1999) defines information science as 'the science that investigates the properties and behaviour of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability'. This is a rather cumbersome definition given that it is difficult to determine the behaviour of information. Miranda Lee Pao (1989:3) seems to have a much better explanation in her attempt when she states that '[t]he processes include the origination, dissemination, collection, organization, storage, retrieval, interpretation, and use of information'. Information science is then 'derived from or related to mathematics, logic, linguistics, psychology, computer technology, operations research, the graphic arts, library science, management, and some other fields'. It is significant to note that information science is not only related to library science as

a discipline, but to a number of other disciplines. It is also argued that both information science and library science draw on other disciplines, to form a meta-discipline (Poole 1985:7). This implies that both library science and information science adopt theoretical frameworks from other disciplines.

Fugmann (1993:1) contends that the fact that 'a definition for the notion of information has been adopted by some information scientists, which has borrowed from another field ... has markedly impeded progress in information science'. As stated earlier, information retrieval research is derived from information science. It is therefore useful to explore the origins of information retrieval.

Information Retrieval Research

Since the 1980s, the cognitive model of information retrieval has probably become the most widely written about (Belkin 1984; Belkin, Brooks & Oddy 1982; Brooks, Daniels & Belkin 1986; Ingwersen 1982). In this particular model, the cognitive state of searchers of information is assumed. The user's conceptualisation of how to formulate a search strategy was investigated often. This in itself is useful. However, even if a user knows how to formulate a search strategy, he or she needs to be familiar with information retrieval systems, for a desired effect (Ingwersen 1992; Saracevic ... et al 1988; Saracevic & Kantor 1988a, 1988b). For the purpose of this chapter the work of two cognitive theorists will be discussed.

Nicholas Belkin

In 1978 Belkin described the problem as follows: 'facilitating the effective communication of desired (stored) information between human generator and human user'. Belkin's (1978:104) statement implies thus that, as a discipline, information science should be dealing with the following issues:

- information in human and cognitive communication systems
- the relationship between information and creator/generator of information
- the relationship between information and user
- effectiveness of information and information transfer

Belkin discusses different kinds of information concepts. His unique contribution is to describe information as a structure: '*the structure of any text which is capable of changing the image structure of a recipient*' (Belkin & Robertson [1976:201] cited in Belkin [1978:80]). This he summarises in the following table.

The communication system of information science (Belkin 1978:81)
GENERATORS → TEXTS ←→ RECIPIENT (a) The Linguistic Level of the System
STATES OF KNOWLEDGE → INFORMATION ←→ ANOMALOUS STATES OF KNOWLEDGE (b) The Cognitive Level of the System

The system can be briefly interpreted as follows.

A potential creator of information decides he or she wants to communicate an aspect of his or her knowledge. This knowledge is adapted to the target group to be reached and creates a text in this way. Belkin describes text as 'a collection of signs purposefully structured by a sender with intention of changing the image-structure of a recipient', and this text becomes part of the wider whole to which receivers have access.

Knowledge gaps identified by the user may be likened to those existing in the mind of somebody wanting to communicate the same thing. The receiver changes this lack of knowledge or Anomalous State of Knowledge (ASK) into a question. This can be used to retrieve the apposite information from the relevant body of knowledge. If the retrieved texts satisfy the user's needs the process is completed. If not, it is repeated with a new ASK (Belkin 1978:80-81).

Peter Ingwersen

Ingwersen perceived in the 1980s that changes were happening in information science, away from an emphasis on the technical aspects towards a more human approach and that more emphasis was being placed on the human aspects of information science such as human cognition and human behaviour (Ingwersen 1992:100-102).

Ingwersen (1992:114) developed four focus areas for information science research:

- infometrics, that is the quantities study of the processes of (written) communication, such as co-citation in bibliometrics
- information management, including evaluation and quality of textual and media-based information retrieval systems
- information (retrieval) systems design
- IR (information retrieval) interaction

He reached the conclusion that more emphasis should be placed on the human-driven communication of information. 'There seems to be a strong overall trend characterised by a move of research interests from access-orientation towards accessibility and use of stored knowledge or knowledge representations' (Ingwersen 1992:114).

His views are important for indexers since he stresses the importance of retrieving information needed by users to better understand the concepts and the world in which they work. It has become much more difficult than before to retrieve the needed information from the huge amount of available potential information.

It is then also important to remember that information science focuses on *stored potential information and the retrieval of the required information from that store* (Ingwersen 1992:116). In this process indexing plays a major role in the design of the indexing systems, the analysis of documents, the assigning of indexing terms and the compilation of the final index.

Indexing thus forms an integral and indispensable part of information science. It provides the necessary tools to bring the user in need of information and the potential source of information together.

Indexing and Related Disciplines

Information science can be considered the overarching or umbrella information discipline. Library science, archival science and museology are then seen as sub-disciplines or part disciplines of information science; information or data processing may also be included as a sub-discipline of information science.

Library Science

In library science, indexing forms part of the sub-field of information organisation and retrieval (IOAR). In IOAR the construction of surrogates, that is information records, is studied. These surrogates include the physical description of the information source, conceptual analysis, designating subject headings and/or classification notations, designing and maintaining the indexing system (a catalogue, a bibliography, an index, etc.) and the systematic study of the effectiveness of the index in terms of user satisfaction and precision criteria among other things.

Information organisation and retrieval is one of the core sub-fields of library science and is one of the areas where the theoretical basis of information science is tested in practice, and refined and improved. Indexing forms one of the core competencies or intellectual capital of a librarian. It is also a generic competency that can be applied in other disciplines as well.

Archival Science and Records Management

The same applies to archival science and records management. The only difference is that the indexing languages and methods may differ. In the former, the principle of provenance (*respect des fonds* or *herkomstbeginsel*) may sometimes be more important than the subject. This does not imply that the subject approach is disregarded or of lesser importance; on the contrary, archivists and records managers do provide this access. The use of computers has made this access even more important and easier to provide.

The various indexes, guides and registers (finding aids) compiled by the records manager or archivist are all indexes. Often these indexes are expanded to include background or contextual information on the origin of the records or archival collection or group. Matters such as when, where, and why the records were created, what is the value of an archives, et cetera, are included. Archivists and records managers can be said to compile enriched indexes, but the basic principles used are the same.

Museology

What is said above about library science, archival science and records management is also true of museology. The basic principles are the same – it is just the form and contents of the index that may differ.

Museums work primarily with objects or artifacts, which can be *naturalia* (natural) or *artificialia* (human-made). These objects should be described as completely as possible physically, and identified. To what species does this spider belong? Additional information

is provided such as where and by whom it was collected, the characteristics of the species, et cetera.

As in the case of archivists, museologists also compile enriched indexes but the principles used are those of indexing.

Linguistics

Linguistics is the discipline that studies *inter alia* the theory of language, its origins, relations between words and graphic signs. How is a sentence constructed? How should words be arranged in a sentence to convey a certain meaning? For indexers, knowledge of linguistics is essential to compiling an indexing language, with a special focus on semantics.

Computer Science and Telecommunication

Computers and telecommunication create the data and information networks such as the Internet that now need human intervention to index and provide access for users. Although attempts are made at computerised indexing, indexing remains basically a human activity. Nevertheless, computer scientists and telecommunication engineers require knowledge of indexing to design user-friendly computerised IR systems. Human indexers are especially needed to help design user-friendly interfaces.

Philosophy

As a discipline, philosophy provides a solid foundation for indexing. Epistemology, the theory of knowledge, explains to us how knowledge is created in society, how an individual gains knowledge and how knowledge is organised. This determines how indexers design an index, designate indexing terms, et cetera. In the case of knowledge, our epistemological position plays a powerful role. This means that the way knowledge is viewed, for instance the truth is either relative or precisely definable, will determine how an index is compiled, headings or search points, and references are decided upon.

It can be stated that indexing, both as activity and as discipline, is related to the use of knowledge and techniques from other disciplines. To a certain extent it can be described as an interdisciplinary or multidisciplinary field of study. Together with communication and information science it may perhaps qualify as an 'interscience'.

What is Indexing All About?

To summarise, indexing is about providing access to potential information to satisfy the needs of users in need of information. To be able to do this, the indexer should describe the information source, a book, journal, video recording, computer file, image, a person and so forth, in order to enable the searcher to be able to trace it. This description can be of the physical item only (descriptive cataloguing), but to make this information source optimally traceable the contents or information in the source should also be described. This is done through the process of conceptual analysis whereby the indexer determines the subject of the source.

After this process, the subject terms are translated into the indexing language (e.g. a controlled vocabulary) and are included in the index. The index itself is edited, presented in a user-friendly format, instructions for its use provided and, if required, continuously updated and improved.

With indexing, a surrogate record of the original potential information source is created. Rowley (1996:158) offers a useful definition of a record when she states that a record 'is the information contained in a database relating to one document or item'. An entry in a manual or online catalogue, which relates to a book or journal article, can therefore be regarded as a record for the book or journal article. To Keenan and Johnston (2000) a record is a 'computing-unit in a computer file usually made up of fields'. In terms of information retrieval, a record is also the 'bibliographic description of an item in the published literature or other recorded material' (Keenan & Johnston 2000:209). When looking for information, the user need only consult these surrogates as the first step in the search process to retrieve possible sources of information. In step two, the user can consult the information source itself. This implies that indexers should describe the information record as accurately as is possible to achieve maximum recall, exhaustivity and relevancy. Recall refers to the number of sources retrieved; exhaustivity implies that all possible information sources are retrieved; relevancy to the closeness of the match between the searcher's request and entities retrieved. Some authors take the definition of recall further, for example recall may be defined thus: '[It is] the proportion of relevant matter retrieved from the total relevant matter in the database. It is a measure of the degree of efficiency of a retrieval process' (Townley 1978:107). Recall also relates to the ability of a system 'to retrieve relevant information' (Keenan & Johnston 2000:208). Relevance, exhaustivity and recall determine the precision of the index. Precision in the context of information retrieval and indexing can be defined thus: '[It is] the proportion of retrieved material relevant to the enquiry. It is a measure of the degree of accuracy of the retrieval process ... [P]recision and ... recall are dependent on the efficiency not on the retrieval program, but of the indexing language used by the system' (Townley 1978:106). Indexers should thus strive to index entities to the highest level of precision.

Indexing fulfils another function as well. By creating indexes, a contribution to the national and international information infrastructures is created. An index created in an isolated community such as Lesotho, for instance, ultimately forms a small but important link in the world's information infrastructure.

The Indexer

What is expected of a competent indexer? Various lists of competencies that good and successful indexers should possess can be compiled. Authors such as Bridgeway Career Development (2001), Lancaster (1998), Wallis (1997) and Chu and O'Brien (1993) all compiled lists. For this contribution a core list is presented, based partly on the above and partly on the author's own experience.

Ideally, a professionally educated and trained indexer should possess the following core competencies:

- A sound general education in information science, which can also include librarianship, archival science, museology, publishing and/or any other academic subject.
- An understanding of knowledge management and related concepts such as data and information. How is information and knowledge generated, produced and regulated in society? Other matters can be the recording, organisation for retrieval of information, the search for information and its use in society.
- A good understanding of the theory of indexing, its role in providing access to information and its communication.
- Language skills (e.g. semantics, syntax).
- Competence in different types of indexing: indexing catalogues, book indexes, databases, manual and automated indexes, and indexing Internet data, for example. A substantial knowledge of indexes to electronic records is becoming increasingly more important.
- Knowledge and information in all forms and formats – books, pictures, musical instruments, Scottish tartans, CD-ROMs, technical reports, et cetera. Ability to answer the question 'What can be indexed?'
- A genuine understanding of the indexing process, that is of conceptual analysis and translation. The indexer should also be able to apply his or her knowledge and successfully complete a number of different indexes, for example to books and electronic materials. The indexer should also demonstrate an understanding of, and ability to use, manual and automated or electronic methods and systems of indexing.
- A qualified indexer should be able to undertake user studies, to evaluate indexes and indexing systems, to select appropriate indexing systems or indexes and to modify and improve existing ones in accordance with user needs and requirements.
- An understanding of the role and use of technology in indexing, including computers, telecommunications and the requirements of the digital library will also be essential abilities for a successful indexer in the digital age.
- The successful completion of a practical component in the form of an internship, practical sessions at book publishers or Internet providers and the compilation of different kinds of index can also be required of qualified indexers and can be a valid learning outcome.
- Entrepreneurship, especially for freelance indexers, quoting fees for indexes, marketing of services, knowledge of accounting, budgeting, time and stress management, human resource management, office technology and administration are also vital skills that should be included in the curricula for the training and education of indexers.
- Indexers should be taught professional and ethical conduct and the distinctive characteristics of their profession. These include matters such as a service orientation, punctuality, confidentiality, acknowledgment to clients of the inability to compile an index, a clash of interests and the like. Although these topics are often seen as peripheral or optional, they are very important components of an education and training programme for prospective indexers.
- Participation in continuing professional education and lifelong learning.

In conclusion, an indexer should be a curious, knowledgeable, meticulous and self-motivated individual, be able to work on his or her own, be creative, professional and a successful entrepreneur. These qualities imply that an indexer is a very special person and the profession of the indexer a worthwhile one.

Capita Selecta of Some Indexing Problems/Issues

Automatic Indexing

Since the use of computers in the compilation and production of books and journals, the idea evolved on using the computer to index the enormous amounts of information being published. Some of the first attempts involved the listing of keywords in the title of a journal. This could be done in context, known as KWIC (keyword in context) indexing, or out of context, known as KWOC (keyword out of context). Although an index is produced in this way, it has a limited value. Insufficient information of the entity is displayed, and the title could be misleading.

Luhn (1959) formulated the KWIC method whereby many points of access are provided for conducting searches. The KWIC index was derived from titles of publications (Lancaster 1998:48). According to Lancaster (1998:48) each keyword in a title 'becomes an entry point and is highlighted in some way, usually by being set off at the centre of a page'. The computer program prevents the rest of the title information from being used as entry points. Some observers found that although the KWIC method was popular, inexpensive and a relatively quick means of the production of indexes, other more sophisticated machine-generated methods superseded them (Foskett 1996:35).

The KWOC method relates to keywords in titles becoming access points out of context. These access points, unlike KWIC terms, are set off in the left-hand margin of the page (Lancaster 1998:49).

Although titles in KWIC and KWOC methods generate a number of entries, recall will be low, while relevance will possibly be high. Foskett (1996:36) argues that with both indexing techniques 'specificity depends on the authors' choice of words, while exhaustivity again depends on how detailed the titles are'. Clearly, although users may have found some entries by searching for a term in KWIC and KWOC indexes, the likelihood of not finding specific entries cannot be ruled out. If you are interested in KWIC and KWOC, the book by Cleveland and Cleveland (see bibliography) might be useful.

Computer scientists eventually realised that indexing has a human component: conceptual analysis, the linking of fields that seem illogical to the computer. Artificial intelligence was used to prepare improved indexes, but with limited success. A combination of human input and computers gives better results, but it still cannot equal a well-prepared index by an experienced indexer.

However computerised tools proved to be more useful, especially to eliminate some of the routine tasks. In this regard word processors proved to be more efficient. These matters receive more attention in chapter 10 which is on automatic indexing.

Web or Internet Indexing

The development of the Internet and the growth of websites containing useful information for users created new challenges for indexers. Special search programs were developed to search the Internet. These are known as search engines – examples are Google, Yahoo and Anansi.

Although these search engines do a relatively satisfactory job, most of them recall a large number of irrelevant references. The process lacks precision and also fails to search the 'hidden web'. Indexers are being compelled to study Internet indexing and to help design search tools for the Internet. See chapter 10 for more information.

Subject Analysis

This will continue to be a problem for indexers since the process is controlled by a number of factors. These include current discourses in society, for instance political correctness as an issue, social conditions (politics, availability of sufficient funds for proper in-depth analysis, forthcoming publications which will create the opportunity for an index), the influence of other disciplines such as semiotics and epistemology, and the growth in the volume of indexable materials or records. Subject analysis also plays an important role in the search for information, as already indicated.

Technology

References to technology featured often in the preceding text. Technology will continue to be important and will influence indexing in many ways. A few examples will suffice. Technology will determine how many records, what kinds of record, the physical forms of the record and how many indexes need to be prepared. On the other hand, technology may give us solutions to some of our indexing problems such as automatic indexing. We cannot predict to what extent software developments such as expert systems can help in (or even do) subject analysis, without or with minimal human input.

Political Economy of Information

Indexing and indexers are all part of a country's information industries. The status, growth and decline of this sector will influence the production and compilation of indexes and ultimately the need for indexers and indexing.

New Knowledge Systems

We are witnessing the growth of non-Western knowledge systems, such as the indigenous knowledge systems of African, Aboriginal or Maori origin. Western scientific knowledge systems are changing under the influence of postmodernism and other philosophical schools. To date, in South Africa, indexers have been accustomed to working mainly with Western traditional knowledge systems, which begs the question how these new developments will be accommodated.

Indexers in a Monetarist World Order

Monetary policies, such as Thatcherism, also pose a threat to indexing. The maxim 'If it does not pay or show a profit, shut it down' will result in indexers perhaps being increasingly called to justify their services and products in financial terms of profit and loss.

Ethics

Last, but not least, ethical issues can create problems for indexers. It is often said that the growth in information creates an equal growth in ethical problems for indexers. Will we be able to deal with all of these ethical issues?

Conclusion

This chapter attempted to put indexing into context as an indispensable part of the whole universe of knowledge and, more specifically, in the field of information organisation and retrieval. The role of indexers and indexing in society was also discussed as well as some of the future problems facing indexing.

BIBLIOGRAPHY

Bates, MJ. 1999. The invisible substrate of information science. *Journal of the American Society for Information Science*, 50(12):1043-1048.

Belkin, NJ. 1978. Information concepts for information science. *Journal of Documentation*, 34(1):55-85.

Belkin, NJ. 1984. Cognitive models and information transfer. *Social Science Information Studies,* 4(2/3):111-129.

Belkin, NJ. 1990. The cognitive viewpoint in information science. *Journal of Information Science*, 16:11-15.

Belkin, NJ, Brooks, HM & Oddy, RN. 1982. Ask for information retrieval. *Journal of Documentation*, 38(2):61-71.

Belkin, NJ & Robertson, SE. 1976. Information science and the phenomenon of information. *Journal of the ASIS*, 27:197-204.

Belton, BK. 2003. *A design foundation for information architecture.* Available: http://eprints.rclis.org/(accessed 4 April 2003).

Bridgeway Career Development. 2001. *Career profile: Indexer: interview by Janet Scarborough with Carolyne G. Weaver....* Available: *http://www.bridgewaycareer.com/profile-indexer.htm* (accessed 4 April 2003).

Brooks, HM, Daniels, PJ & Belkin, NJ. 1986. Research on information interaction and intelligent information provision mechanisms. *Journal of Information Science*, 12(1/2):37-44.

Chu, CM & O'Brien, A. 1993. Subject analysis: the critical first stage in indexing. *Journal of Information Science*, 19:439-454.

Cleveland, DB & Cleveland, AD. 1990. *Introduction to indexing and abstracting.* 2nd ed. Englewood, Colo: Libraries Unlimited.

De Beer, CS. 1998. 'n Skets van die hedendaagse inligtinglandskap. *Mousaion 3*, 16(1):74-103.

Foskett, AC. 1996. *The subject approach to information.* London: Library Association Publishing.

Fourie, I. 2002. How can we take a socio-cognitive approach in teaching indexing and abstracting? *The Indexer*, 23(2):83-84.

Fugmann, R. 1993. *Subject analysis and indexing.* Frankfurt: Indeks Verlag.

Indexing. Available: http://www.author.co.uk/indexing.htm(accessed 4 April 2003).

Ingwersen, P. 1982. Search procedures in the library – analyzed from the cognitive point of view. *Journal of Documentation,* 38(3):165-191.

Ingwersen, P. 1992. Information and information science in context. *Libri*, 42(2):99-135.

Keenan S & Johnston. 2000. *Concise dictionary of Library and Information Science.* London: Bowker-Saur.

Lancaster, FW. 1998. *Indexing and abstracting in theory and practice.* 2nd ed. Champaign, Ill: University of Illinois, Graduate School of Library and Information Science.

Lancaster, FW & Warner, A. 2001. *Intelligent technologies in library and information science applications.* Medford, NJ: Published for the American Society for Information Science and Technology by Information Today.

Luhn, HP. 1959. *Keyword-in-context index for technical literature (KWIC Index).* Yorktown Heights, New York: IBM Research Centre.

New Encyclopaedia Britannica. 1990. Vol 21. *Information processing and information systems.* Chicago: Encyclopaedia Britannica.

Pao, ML. 1989. *Concepts of information retrieval.* Englewood, Colo: Libraries Unlimited.

Poole, H. 1985. *Theories of the middle range.* Norwood, NJ: Ablex.

Rowley, J. 1996. *The basics of Information Systems.* London: Library Association Publishing.

Saracevic, T ... et al. 1988. A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science*, 39(3):161-176.

Saracevic, T & Kantor, P. 1988a. A study of information seeking and retrieving. II. Users, questions, and effectiveness. *Journal of the American Society for Information Science*, 39(3):177-196.

Saracevic, T & Kantor, P. 1988b. A study of information seeking and retrieving. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science*, 39(3):197-216.

Sharp, JR. 1968. *Some fundamentals of information retrieval.* London: Deutsch.

Schrader, AM. 1984. In search of a name: information science and its conceptual antecedents. *Library and Information Science Research,* 8(3):227-242.

Taylor, AG. 1999. *The organization of information.* Englewood, colorado: Libraries Unlimited.

Townley, HM. 1978. *Systems analysis for information retrieval.* London: Deutsch.

Universal dictionary. 1987. London: Readers Digest.

Wallis, E. 1997. Indexing as a professional activity. *The Indexer,* 20(4):189-191.

Weinberg, BH. 1993. American Society of Indexers: history, activities and relationships to ASIS. *Bulletin of the American Society for Information Scientists,* February-March:23-24.

Wersig, G. 1997. Information theory, in *International encyclopaedia of library and information science,* [edited by] DJ Feather and P Sturgess. London: Routledge:220-227.

3 CONCEPTUALISATION: WHAT IS IT ALL ABOUT?

Ina Fourie

Abstract

Conceptualisation and aboutness are key concepts in IR (information retrieval) and the preparation of representations of documents or other entities. The complexities of conceptualisation, guidelines on conceptualisation, different views of conceptualisation, and the meaning of aboutness are considered. In addition the socio-cognitive approach to document representation is also briefly mentioned.

Introduction

IR (information retrieval) is concerned with the processes involved in the representation, storage, searching and finding of information that are relevant to a requirement for information as desired by a human user. We find such representations in information retrieval systems (IRS) such as databases, library catalogues, intranets, and portals. We can even find representations in full-text databases and the index at the back of a book. We create representations to support effective and efficient IR. In databases we refer to such representations as records. We can also refer to them as surrogates. 'The central problem of IR is how to represent documents for retrieval. The most intricate or carefully designed algorithm cannot compensate for inappropriately represented documents' (Blair 1990:vii).

When we create records for entities such as documents, museum objects, archival records, or websites, we describe them as

- Physical entities (i.e. things with characteristics such as an author's name, a title, an ID number, an URL) that can be used to identify the entity. Such characteristics are used to trace the entity when needed. One needs such characteristics when tracing an article to photocopy or to download it from a full-text electronic journal database, or to order a book from a publisher.
- Entities with intellectual content (i.e. entities containing information that might be desired by someone hoping to solve a problem).

Ingwersen (1992:50-54) refers to the above as *isness* (the description of the physical entity) and *aboutness*. 'What is this document or text about?' One may also refer to aboutness as topical representation (i.e. asking which topics are covered by the document). The title, abstract, indexing terms (also called descriptors), classification notations, identifiers and subject headings all reflect the *aboutness* of an entity. They are the end products of conceptualisation (also referred to as conceptual analysis).

There are many reasons why representation is so difficult and complex. The three most important are these complexities:

- conceptualisation (deciding what an entity is about)
- selection of the concepts that are worth mentioning in a representation (exhaustivity[1])
- the translation of these concepts into words (the issues of vocabulary and specificity[2]) (discussed in chapter 6 on verbal subject description).

Conceptualisation and entity representation are not just important to indexers and abstractors, but also to those involved with portal and intranet design, and information and knowledge management.

The focus of this chapter is on conceptualisation. In the following sections, the author focuses on

- how conceptualisation fits into the process of entity representation
- the concept of aboutness
- guidelines on conceptualisation
- different approaches to conceptualisation
- the complexity of conceptualisation

How Conceptualisation Fits into the Process of Entity Representation

Conceptualisation is necessary to create entity representations that are stored in an IRS to be retrieved by a person desiring information to fill a gap between what one knows and what one needs to know to solve a problem.

The quality of subject descriptions such as indexing, abstracting and classification depends on the conceptual analysis that precedes it. The effective retrieval of information is only possible if the input at the stage of conceptual analysis and subject description, which precedes retrieval, is of a high standard. This is only possible if the techniques of conceptual analysis ensure that indexers index useful concepts from the information source (e.g. a document) as accurately and consistently (inter-consistently and intra-consistently) as possible.

Concept of Aboutness

'Conceptual analysis, first and foremost, involves deciding what a document is about – that is, what it covers' (Lancaster 2003:9). Many authors have speculated on 'aboutness' (e.g. Bruza, Song & Wong 2000; Hjørland 2001). It has also been associated with terms such

1 Exhaustivity refers to the degree to which the subject content of an entity is reflected in the entries. It also indicates the degree to which information about a subject can be retrieved.

2 Specificity of indexing refers to the degree to which an index entry matches the exact meaning of the subject concept (Cleveland & Cleveland 2001:259), or the vocabulary used in the entity.

as coverage, subject, themes, domains and fields, and subject access points (Hjørland & Nielsen 2001). The question even arises whether it is really important to know the meaning of aboutness. Lancaster (1991:11) remarks: 'However, if one must reach agreement on the precise definition of terms before pursuing any task one is unlikely to accomplish much – in indexing or any other activity'. According to Lancaster (2003:13) it is sufficient to recognise that an entity is of interest to a particular user group, and to translate these into concepts that will support IR.

What is a Concept?

In the field of entity representation the following explanations of a concept have been offered:

- a topic discussed by the author
- themes
- units of thought
- things or matters discussed
- philosophical content
- focus of a work
- individual subjects, their parts and the relationships between the parts
- central themes on which the author focuses his or her attention and energy

There is, however, more to conceptualisation than just selecting concepts based on the content of a document or other entity. One needs to take a deeper look at the hidden content, and the importance of the content to the potential user group.

Guidelines on Conceptualisation

The process of conceptualisation is general: it is not influenced by the requirements for the different kinds of translation into words such as indexing terms, abstracts, classification notations or subject headings. According to Lancaster (2003:15) conceptual analysis means 'nothing more than identifying the topics that are discussed or otherwise represented in a document'.

There are many formal publications on the rules for both the physical descriptions and descriptions of the intellectual content (e.g. *Recommendations for examining documents, determining their subjects and selecting indexing terms* 1984). Standards are also discussed by Anderson (1994).

Such publications also offer guidelines on conceptualisation. Guidelines for the conceptualisation (be it for writing an abstract or assigning indexing terms) of an article will for example include the following:

- Read through the document (e.g. article) and get an idea of the subject matter (what it is about). Make notes on the main ideas and other important information. Especially note the title, headings, subheadings, introduction, closing paragraph or summary, tables, figures, graphs, first paragraph of each main heading, author's abstract (if available), findings, text of the article (special attention should be given to words in bold or italics, graphs, figures, tables, etc.), references for a document (e.g. bibliography, sources consulted) (This applies only in problematic cases.)
- Other guidelines: the information included in your abstract should be accurate. If the abstract is for a journal article on a research project, one should for example include the following: aims and objectives of the research, scope, research methodology, results, findings and recommendations. If it is not a research article, consider the content of the article. For indexing terms, there should be sufficient information to warrant the use of the indexing term. However, if it is a new concept or idea, it should be mentioned, even if briefly.
- Other: There are also guidelines on the length of a description informing one what the entity is about, the style of abstracting (e.g. one paragraph only) and the concise use of language (covering the maximum amount of information by using as few words as possible, avoiding unnecessary words like 'this article' and 'this report').

Some of the above applies more to the translation phase, but is also relevant to conceptualisation.

These considerations seem very simple. Why is conceptualisation then so complex?

It is also worth noting Hjørland's (2001:774) adaptation of the guidelines offered by Wilson on how to determine a document's content:

- Determine the author's intentions and purpose in writing the document.
- Read the document and weigh the relative dominance and subordination of different elements covered in the document.
- Group or count the concepts mentioned in the document.
- Develop a set of rules to determine what is essential in the document.

Different Views of Conceptualisation

Many people have theorised on aboutness and conceptualisation. Getting down to more practical details, the following views of Albrechtsen (1993) on conceptualisation are considered. (The terms used here are the terms identified by Albrechtsen. However, I added my own explanations to simplify the discussion.)

Simplistic Approach

Subjects covered in entities are regarded as absolute objective entities that can be derived from the content of entities such as documents (e.g. as in automatic indexing and abstracting): one merely picks the topics from the words in the text (e.g. the title and main headings). It is a very pragmatic approach, not too time-consuming and much easier to apply than the other approaches. But it means no value is added (e.g. in linking the content to specific problems that it might solve) and that users' needs and the potential usefulness of the entity are not considered. Looking at this chapter, for example, one could choose the following concepts:

Aboutness, conceptualisation, simplistic approach, content-oriented approach, socio-cognitive approach.

The simplistic as well as the content-oriented approach may be compared to Soergel's (1985:227) entity-oriented indexing.

Content-oriented Approach

The content-oriented approach includes an interpretation of an entity's contents that involves the identification of topics or subjects that are not explicitly stated in the text. The person responsible for the conceptualisation discerns implied concepts. An article on the University of South Africa, for example, also implies distance education. Identification of implied concepts requires prior knowledge of the subject matter. Looking at this chapter again, one could for example choose the following concepts, in addition to those identified for the simplistic approach:

Information organisation, subject analysis, indexing, document representation

Although the content-oriented approach takes a deeper look at the content of an entity, it still does not consider the users' needs, or the potential use of the information contained in the entity.

Requirements-oriented Approach

The requirements-oriented approach takes into account possible requests from users and also considers contextual frameworks. Albrechtsen (1993:222) explains that the person responsible for the subject analysis should ask '[H]ow should I make this document, or this particular part of it, visible to potential users? What terms should I use to convey its knowledge to those interested?' An entity is analysed to predict its potential usefulness. This can be compared to Soergel's (1985:230) request-oriented approach.

The advantage of the requirements-oriented approach is that it can assist both in solving problems experienced by users, and in the transfer of knowledge. It is more time-consuming than the other approaches and requires a sound knowledge of the users, their needs and the problems they need to solve. Once again, looking at this chapter, one could

for example select the following concepts, in addition to the concepts identified for the other two approaches:

What types of conceptualisation can be distinguished?
- Conceptualisation views

Which approaches to conceptualisation allow for users' needs?
- User-centred indexing

The socio-cognitive view can be linked to the requirements-oriented approach. Hjørland and Albrechtsen (1995) initially refer to this as the 'domain analytical approach'. Fidel (1994) writes about user-centred indexing.

Socio-cognitive View

The question arises whether we should consider the users as individuals, or as people belonging to a larger group, sharing similar information needs and vocabulary. The socio-cognitive approach seems to stress the latter (Hjørland 2002a, 2002b). According to Jacob and Shaw (1998:142) the socio-cognitive approach shifts attention from individual knowledge structures to the discourse domains of knowledge-producing, knowledge-sharing and knowledge-consuming communities.

A simplification of the socio-cognitive approach will read as follows: people live, work, study and communicate in a social environment. They form part of a group of people who talk to one another (a discourse community). They are influenced by their culture, historical developments, political ideologies, economic situation, education and work environments, and so on. This can be referred to as a socio-cultural environment that applies to children, students, academics from different disciplines, professionals from different sectors, et cetera. These people communicate with one another. They share information, ideas, differences and motivation. They learn from the 'history' of the particular group and the inherent culture. They share a common language (to some extent) of concepts, theories, expressions, methods, et cetera. They use this language (vocabulary) when

- searching for information (i.e. interacting with document representations such as indexing terms and abstracts)
- communicating information (e.g. informal discussions, conference papers, articles, books)
- producing or describing information (e.g. news reports, lectures, adverts, building plans, software design)
- using information (e.g. interpreting published documents in terms and concepts that they can use for their productions)

The socio-cognitive approach applies to conceptualisation as well as translation.

In their review article Jacob and Shaw (1998) refer to a number of research projects that support the importance of considering specific user groups. Wilson (cited by Jacob & Shaw [1998:150]) for example found that social workers' document representations should focus on problems they actually face. In his research, Swift (cited by Jacob & Shaw [1998:150])

found that social scientists need multiple document representations for all documents (e.g. theoretical focus, research methods, research paradigms). Tibbo (1992, 1994) found that specialised guidelines are necessary for indexing and abstracting in disciplines from the humanities, such as history.

Hjørland and Christensen (2002) also link the socio-cognitive approach to information seeking.

In IR, users interact with document representations through their IRS and its search engine. If we can learn more about the vocabulary applying to a particular group, and their needs, it might be possible to generalise our expectations for the group, and to allow for this in entity representations. Such insights will influence conceptualisation, selection of concepts that are worthy to include in a representation, and the actual translation of the concepts into vocabulary.

We can apply the above views of conceptualisation to this chapter as follows.

Different facets of conceptualisation	Where/how to look
Simplistic approach Which topics are obviously discussed in the text? Which of these topics are important enough to include?	Title, introduction, author's abstract, main headings, subheadings, etc.
Content-oriented approach Are there any topics implied in the discussion?	Text, and indexer's experience and interpretation
Requirements-oriented approach How can the readers/users benefit from the discussion? Which questions might be answered by the text? What might they like to know? Which problems can be solved? Are only topics mentioned, but of importance to users (e.g. a new product on the market, new research with a radio isotope)?	Sound knowledge of potential users, their background, knowledge and needs. Can obtain information through interviews (focus group or individual), questionnaires, content analysis of publications and conference papers, etc.
Socio-cognitive approach How can knowledge of the potential readers/users as part of a discourse community be used to improve access to the entity (chapter)? · Which of these things would be of particular interest to the users? · Are there perhaps different users with different information needs? · Are there specific things that the users might be interested in? · Do the users have different levels of experience and subject knowledge?	Sound knowledge of potential user group. Can obtain information through interviews (focus group or individual), questionnaires, content analysis of publications and conference papers, etc.

Complexity of Conceptualisation

Even if we support the simplistic view of conceptualisation, it is more complex than just picking concepts from the content of an entity.

- There is never just one way of looking at the intellectual content of an entity. People will notice different things from the same entity. This is referred to as inter-indexer inconsistency.
- A person will see different things when looking at the intellectual content of an entity at different points in time. This is called intra-indexer inconsistency.

Conceptualisation is also influenced by for example the knowledge, background and experiences of the person responsible for the conceptualisation. This is referred to as a cognitive framework and includes knowledge of the document's subject, experience in conceptualisation, insight into both the indexer and the potential user.

Conclusion

No matter how much is written on conceptualisation and aboutness in document representation, it will always be a difficult and complex process. It requires reflection on what is understood by conceptualisation, and how it impacts on indexing and abstracting, and so on, thereby gradually improving the indexer's understanding of conceptualisation.

BIBLIOGRAPHY

Albrechtsen, H. 1993. Subject analysis and indexing: from automated indexing to domain analysis. *The Indexer*, 18(4):219-224.

Anderson, JD. 1994. Standards for indexing: revising the American National Standards Guidelines Z39.4. *Journal of the American Society for Information Science*, 45(8):628-636.

Blair, D. 1990. *Language and representation in IR*. Oxford: Elsevier.

Bruza, PD ... et al. .... 2000. Aboutness from a commonsense perspective. *Journal of the American Society for Information Science*, 51:1090-1105.

Cleveland, DB & Cleveland, AD. 2001. *Introduction to indexing and abstracting*. 3rd ed. Englewood, Colo: Libraries Unlimited.

Fidel, R. 1994. User-centred indexing. *Journal of the American Society for Information Science*, 45(8):572-576.

Hjørland, B. 2001. Towards a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance. *Journal of the American Society for Information Science and Technology,* 52:774-778.

Hjørland, B & Nielsen, LK. 2001. Subject access points in electronic retrieval. *Annual Review of Information Science and Technology*, 35:249-298.

Hjørland, B. 2002a. Domain analysis in information science: eleven approaches – traditional as well as innovative. *Journal of Documentation*, 58(4):422-462.

Hjørland, B. 2002b. Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Technology,* 53(4):257-270.

Hjørland, B & Albrechtsen, H. 1995. Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science*, 46(6):400-425.

Hjørland, B & Christensen, FS. 2002. Work task and socio-cognitive relevance: a specific example. *Journal of the American Society for Information Science and Technology,* 53(11):960-965.

Ingwersen, P. 1992. *IR interaction.* London: Taylor Graham.

Jacob, EK & Shaw, D. 1998. Socio-cognitive perspectives on representation. *Annual Review of Information Science and Technology,* 33:131-185.

Lancaster, FW. 1991. *Indexing and abstracting in theory and practice.* Illinois: University of Illinois.

Lancaster, FW. 2003. *Indexing and abstracting in theory and practice.* 3rd ed. London: Facet Publishing.

Recommendations for examining documents, determining their subjects and selecting indexing terms. 1984. London: British Standards Institution. (BS 6529:1964).

Soergel, D. 1985. *Organizing information: principles of data base and information retrieval systems.* San Diego: Academic Press.

Tibbo, HR. 1992. Abstracting across the disciplines: a content analysis of abstracts from the natural sciences, the social sciences, and the humanities with implications for abstracting standards and online IR. *Library & Information Science Research*, 14:31-56.

Tibbo, H.R. 1994. Indexing for the humanities. *Journal of the American Society for Information Science*, 45(8):607-619.

4 CHARACTERISTICS AND EVALUATION OF A GOOD INDEX

Anna-Marie Arnold

Abstract

The evaluation of a completed index is a vital step towards the mastering of quality and competence in the production of indexes. However, the fact that there is no single and absolutely correct way to construct an index makes index evaluation problematic. A brief review of the literature on the theoretical base and practice concerning the evaluation of indexes is presented. Valuable guidelines are available for the process of indexing and the presentation of an index. The indexing process can be influenced by authors, publishers, regulators, subjects and users, but basic requirements should be met by the professional indexer. Indexing evaluation can be beneficial in a number of ways. It can benefit the indexer as a means of self-assessment, assist in the assessment of training courses in indexing, be applied by societies for indexers as a means to offer recognition to the competence of members of the society, help publishers to identify competent indexers for the indexing of publications, and serve to raise the standard of indexing practice and professionalism.

Introduction

The evaluation of a completed task should never be underestimated. This chapter explores the theoretical base and practice of the evaluation of indexes. A review of the literature on indexing shows that there are many publications dealing with practical descriptions on the practice of indexing but very few on the theory of indexing based on empirical studies, such as the evaluation of indexes. The evaluation of an index is complicated, mainly because of the individual nature of the indexing process. Despite a long tradition in history, the process of indexing is still not well understood and, according to Kascus (1996:254), the field of indexing lacks a generally accepted theoretical base.

Indexes are compiled as vital instruments in the efficient and effective retrieval of information. At a time when access to information is vital to the individual in modern society, it is necessary to evaluate this important type of instrument which is created to assist the information seeker in the process of information retrieval (IR).

Definitions and Functions of an Index

An index is compiled to serve as an instrument in the retrieval of information. According to the British Standard (BS 3700 1988) an *index* is '[a] systematic arrangement of entries designed to enable users to locate information in a document'. The document 'may be a book,

an issue of a periodical (magazine, journal), an audiotape, a film, a computer disk or any other information-carrying artefact whether printed or not' (Society of Indexers 1991:TAB9/91).

The *Chicago manual of style* (1982:512) explains that an index is not merely a simple list or table of subjects. This viewpoint is supported by the American Society of Indexers (ASI) who define an index as follows: '[A]n index is not an outline, nor is it a concordance. It is an intelligently compiled list of topics covered in the work, prepared with the reader's needs in mind' (American Society of Indexers 2001b). Indexes are important pointers to information and may be positioned as an integral part of their system.

'[An index is] a systematic guide to items contained in, or concepts derived from, a collection. These items or derived concepts are represented by entries in a known or stated searchable order, such as alphabetical, chronological, or numerical' (Kascus 1996:255).

Characteristics of an Index

The characteristics of an index may be summed up under the following points (Borko & Bernier 1978:4):

- It facilitates the retrieval of a reference to a specific item or subject.
- A subject can be located within the text of a publication despite the sequence followed in the publication by means of an index.
- An index discloses relationships.
- It discloses omissions.
- The reader is guided to find answers to questions and make discoveries.
- It provides a comprehensive overview of a subject field.
- It provides nomenclature guidance.

Forrester (1995:253) explains that a good index should allow the user to move around a text and follow up lines of enquiry and investigation on a subject. The index needs to enable the user to find information needed or to recall half-remembered passages (Society of Indexers 1991:TAB9/91). This is achieved by the *search terms* provided in the index. The search terms may be 'single words, phrases, abbreviations, acronyms, dates, names, or any other indexable elements or concepts' (Society of Indexers 1991:TAB9/91). A document may also need multiple indexes for different classes of headings, such as an index to common names and another to the scientific names or terminologies.

O'Connor (1996:9) defines the *process of indexing* and links indexing to abstracting as vital for the retrieval of information. His definition reads as follows: 'Indexing points to areas of likely utility, while abstracting provides smaller, secondary documents for inspection. Indexing and abstracting of some sort are absolute necessities for navigating through the sea of information in which we find ourselves today. However, they must be accomplished in ways hospitable to and compatible with those who make use of them' (O'Connor 1996:9). The purpose of the abstract is very different from that of the index. The abstractor represents the arguments and contents of the document in continuous prose form, while the indexer reads the text and distinguishes between the relevant and peripheral information (Farrow 1995:243).

Different Types of Index

There are a number of different types of index:

- Subject and author indexes. These indexes are the ones used most frequently by users.
- Citation indexes. Citation indexes are used to locate earlier references and more recent works that refer to earlier works (Borko & Bernier 1978:149).
- Word indexes. A KWIC index (keyword-in-context for technical literature) is a word index derived from the titles of articles. A KWOC index (keyword-out-of-context) is a word index derived from the full text of the article. A third type of word index is the permuterm index which is a permuted (term-pair or revolved) index, or other special indexes.
- Luhn stresses that a distinction must be made between the indexes that are machine generated and those that are prepared by manually created. He calls the machine-generated index a dissemination index, and the other a retrieval index as it is structured (Kascus 1996:250). The evaluation of the manually compiled subject index forms the primary focus of this chapter, but some reference is also made to automatic or machine indexing.
- Concordances. Concordances are full text word indexes to all the principal words in a single text. Concordances are designed to guide researchers to particular words and their contexts (Borko & Bernier 1978:182).
- Special indexes. These include thematic indexes devoted to a specific theme: ring indexes for chemical compounds, molecular-formula indexes for the names of organic compounds, taxonomic indexes on various life forms, numerical indexes such as a patent-number index, classified indexes used to bring items together in categories (e.g. Medical Subject Headings, which brings all related subject headings together into logical groups) (Borko & Bernier 1978:197).

The index can be viewed as part of an IR system. Figure 1 illustrates the process of IR, with the user's query (Q) for information in relation to the instruments making the retrieval (R) of information possible. The index forms part of the system.

Approaches to the Evaluation of an Index

The evaluation of an index is not a simple task to perform. A brief review of evaluative studies described in the literature shows that the evaluation of an index is both complex and problematic. The main reason is that as there are no set rules guiding the construction of an index. The *Chicago manual of style* states that 'there is no single and absolutely correct way to construct ... an index' (1993:17.53). An index can, therefore, be constructed in more than one way.

Initial studies done in the past identified various criteria as valid for application in the evaluation of an index. These studies aimed to

- assess the achievement of efficient and effective retrieval of information in the index

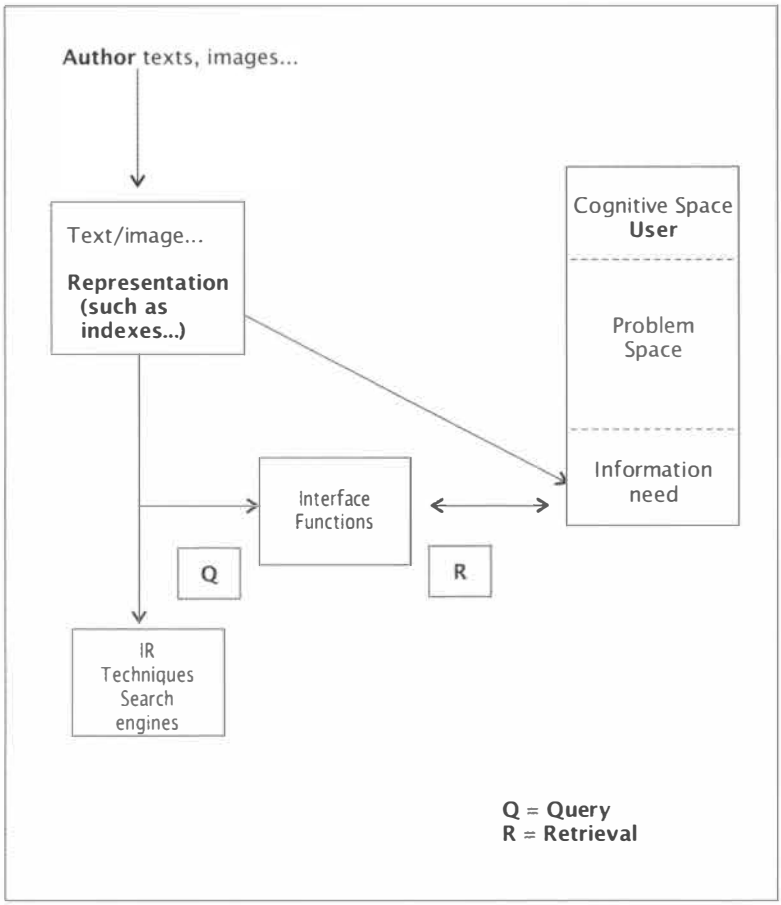


Figure 1: *Indexes as Part of the System and IR Process with the Reader's Needs in Mind* (Ingwersen 1992:55)

- pay more attention to the user of the index in the evaluation of indexes (as in later studies)
- focus on retrieval problems related to automatic indexing (as in more recent approaches)
- consider practical issues concerning the layout, format, framework, consistency, recall, usability and depth of representation

The discipline which is specifically concerned with the study of indexing and abstracting is known as the 'theory of the organization of information' (O'Connor 1996:4,5). This discipline involves itself with fundamental concepts such as information, aboutness, relevance and closeness of meaning. Other matters that fall within this discipline include basic design concepts and automated systems techniques. Within this field the evaluation of indexing is applied by looking at system performance and user satisfaction. The discipline also considers advanced design principles of indexing, and knowledge representation.

The evaluation of an index requires an understanding of the processes involved in the production of the index. The process of indexing in manually constructed indexing is entirely different from that of machine-generated indexing. A distinction needs to be made between describing the process of manually created indexing and the process of machine generated indexing. The design principles and format of representation of these two indexing techniques are also very different.

Manual indexing

The process of manual indexing can be analysed in a number of steps (Wormell 1992:75). These steps are described briefly below.

Extracting Content-bearing Words from the Text

The extracting process is complex because it is situational and subjective. The indexer selects indexing terms to describe the document or image that he or she is indexing. No two indexers will interpret the same text in the same way. The text is about a particular subject area. It is the task of the indexer to interpret the *aboutness* of the text into indexing terms that would represent the subject area. The same indexer may also extract different words or concepts from the text at different times. Furthermore, the concepts do not appear as separate entities in the text, but are always in relation to other terms or concepts. The cognitive process during which the indexer decides to extract content-bearing words from the text is not formally structured. This step in the process of indexing contributes to the unique qualities of the index compiled.

The *aboutness* in the process of indexing can be explained as follows: in the process the author delivers a text or image which is 'about' a particular subject matter (not one specific concept or term); in a process of analysis and interpretation the indexer compiles the index; this index is used by the user who is looking for information 'about' the subject matter which is represented in the index (Ingwersen 1992:55). The elements involved in the analysis are shown in Figure 2.

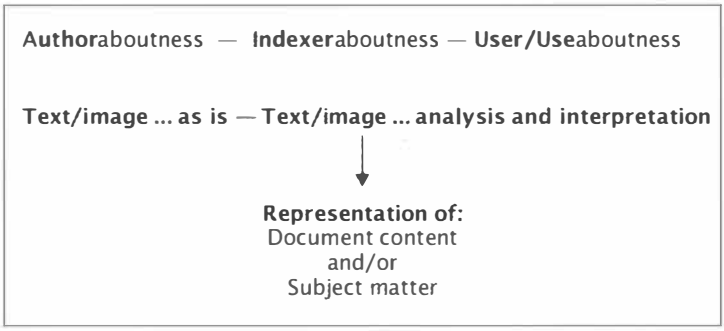


Figure 2: *Elements Involved in the Indexing Process* (Ingwersen 1992:55)

Related Terms and the Original Text

The indexer may consider it necessary to add to or change the vocabulary used in the text while indexing. The terms used in the index are selected by the indexer according to his or her opinion and rational decision.

Expression of the Selected Concepts, if Necessary, in the Language of the Given Indexing Systems

The index needs to be done according to the requirements set out by the indexing system.

Forming the Index (an Ordered List of Terms)

The index is structured and the terms are arranged in an order which will be user-friendly, usually alphabetically, numerically or alphanumerically. The structure or skeleton of the index is constituted by the main entries and sub-entries selected by the indexer. The indexer follows certain formatting rules, decides on the run-on style of entering main entries, or indented style, et cetera.

Examples of entries in an index:

Msimang, Mabaso 19, 75, 90
character of 30
honours received 32-33

Construction of Cross-references

Cross-references are entered by the indexer who groups certain entries in accordance with his or her own judgment. The indexer may also decide to use double-entering of certain terms.

Difficulties Identified in Manual Indexing

Some known difficulties may occur during the process of compiling a manual index. A typical difficulty is that of indexer *consistency*. Research suggests that indexers may have a consistency of between 10 to 80 per cent in areas such as the choice of indexing terms, selection of search strategies, and judgment of relevance (Wormell 1992:80). Difficulties are also experienced in situations where few indexing terms are used because this often affects the level of retrieval performance (Ingwersen 1998). The number of indexing terms assigned in the index would roughly correspond to the exhaustivity of the index. Exhaustive indexing implies that the indexer will use enough terms to cover the subject matter of the document (Lancaster 1991:22).

Automated or Machine-generated Indexing

The procedures for automated or machine indexing are based on the following automated systems techniques (O'Connor 1996:5):

- associated search techniques
- clustering
- automatic extraction
- full text retrieval
- generic algorithms

Difficulties Identified in Automated or Machine-Generated Indexing

The meaning or aboutness of the text is not represented in the automated index. The rules for the extraction of indexing terms require a certain level of manual involvement. This means that a certain level of human involvement is required in the development of the automated indexing system. Automated indexing may also require that the relationship between concepts needs to be pre-defined for automated extraction. The complexities of automated indexing also differs from one language to another and better results are possible in languages which are morphologically more complex than English. The retrieval of documents often requires query modifications and more than one search with new relevance assessments by the user. Other things which may influence the level of IR include the use of synonyms, variant forms of the same word stem, the possibility of an infrequent word still referring to a significant concept, and users looking for concepts not necessarily intended by the user (O'Connor 1996:116). Machine indexing has stimulated and positively influenced the practice of indexing. Kascus (1996:255) argues that machine indexing will remain 'an aid to the process and not a substitute for it'. She reasons that artificial intelligence and expert systems could in the future improve our understanding of the indexing process, which may advance the theory and practice of indexing.

Evaluating an Index

The complexity of the indexing process was recognised with the advent of the computer and the practice of machine indexing in the early 1950s. The development of computers introduced new approaches to the process of indexing. At that time H. P. Luhn recognised the potential of computers for the retrieval of information and for machine indexing. Experiments in indexing evaluation emerged during this period with the Cranfield Projects and other similar projects. During the 1950s, 1960s and early 1970s researchers agreed that there was no 'theoretical base for indexing and also no means to evaluate it'.

Examples of evaluative studies conducted to assess the performance of automatic and/or manually based indexing (with the performance attributes, identified as important for the assessment, printed in *italics*):

- Cranfield Projects (1960-) – under C. Cleverdon – compared four indexing languages (Universal Decimal Classification; faceted classification; coordinate indexing uniterms,

and alphabetic subject headings) looking at *recall* (the proportion of relevant documents retrieved) and precision (the proportion of retrieved documents that are relevant) (Vickery 1973:93).

- Karen Sparck Jones (1979) did research on the relationship between the coverage (exhaustivity) of indexing and *precision* (accuracy) of the indexing language.
- Swanson (1960) recognised the difficulty faced by fully automatic indexing in order to achieve *consistency.*
- Fairthorne observed that the *testing* of indexes could not be considered evaluative studies.
- Bloomfield made a careful distinction between indexing for machine, for machine searching, and indexing for printed indexes, for human searching. For subject indexing Bloomfield lists the following six characteristics of subject indexes: (1) *breadth* of vocabulary (number of different index terms in the subject index), (2) *depth* of indexing (number of index terms assigned to a document to describe its contents), (3) use of *general or specific* terms, (4) use of 'see' and 'see also' references, (5) indexing *format* and (6) inclusion of *titles or other qualifying phrases.* Only the first three listed can be evaluated quantitatively, while the other three can only be evaluated subjectively.
- Artandi identified very similar characteristics to those of Bloomfield. He describes an index's characteristics as follows: *specificity* and *size* of the index language, the network of *relationships* that exist between terms, the *exhaustivity* of the indexing, and the arrangement and *physical characteristics* of the file.
- Lipetz emphasised the nature of indexes based on the decisions on what should be *included*, and how much, and its *usefulness.*
- Lancaster evaluated the performance of indexes from the perspective of the user in terms of *coverage* of the literature, *recall ratio* (ability to yield relevant information), *precision ratio* (amount of effort required to find relevant information) and *novelty factor* (ability to uncover relevant information unknown to the user).
- Calvert (1996:74) observed that the existing *standards* for indexing practice serve mainly as recommendations for indexing and are not intended to promote consistency in the compilation of indexes.

The first major work on indexing was written in 1879 by HB Wheatley, entitled *What is an index?* Since then, much has been published on practical descriptions of indexing, but studies done on indexing are far less empirical (Kascus 1996:245,246). Kascus (1996:254) concludes his chapter on the theory and practice of indexing by stating that it is difficult to 'reach any consensus on what are the essential criteria to be used in measuring performance'.

After reviewing articles on evaluative studies, it appears that the practice of evaluating an index is not only difficult and problematic but may also be considered from various points of view. The *ASI's Indexing evaluation checklist* (American Society of Indexers 2001) sees the evaluation from the point of view of the user of the index, while Hazel Bell (1996:3) takes a wider view. She raises the question 'Whom should we aim to please?' in her article under that title. She refers to five different groups or classes of people to

be 'pleased', each with their own expectations. These classes include the users, subjects, authors, publishers and regulators.

Expectations from Various Viewpoints

User's Point of View

There are many different categories of users, such as adult readers, children, and subject specialists like persons in the medical or other professions or technical fields. Knowledge of the potential user group serves to direct the indexer to include terms that would be considered relevant to the intended user group. The indexer should know who the target audience of a publication is and be guided by this knowledge in the process of indexing. The indexing terms should be appropriate to the intended audience or reader of the publication (American Society of Indexers 2001b:1).

The main user criteria are those of *relevance* and *pertinence*. The user is guided in the index to the required information and also related information. Olason (2000:91) discusses a series of usability studies on indexes done to establish a set of user requirements that indexes need to satisfy. According to the research the indexes used for the test needed to improve their usability (relevance) and quality. The ASI requires that the main headings are relevant to the needs of the reader, and that they are 'pertinent, specific, comprehensive[,] ... not too general ... or too narrow'. The criteria listed for medical indexes include accuracy, thorough analysis, completeness/comprehensiveness and usability (Wyman 1999:126).

However Hans Wellish (Bell 1996:3) warns: 'It has never been the purpose of an index to tell the reader in advance ... what he will find on the page indicated. Index entries are not meant to be mini-encyclopaedias.'

The indexer also needs to assist the user of the index by explaining the conventions applied in the index, thus facilitating the use of the index. This is done in an introductory paragraph to the index. Dixon explains that indexes for children's books should also be provided with an introductory paragraph on the arrangement of the indexing terms. Dixon (1996:8) explains: 'Indexes in children's books need a good explanatory introduction, an indication of where pictures and major references occur in the text, and a large capital letter to indicate the beginning of a new alphabetical division within the index. Subheadings are necessary to avoid strings of references, but perhaps better avoided in books for under eleven year olds.'

The special index requires a structure which reflects the content of the book. The index should add to the value of the book. For the medical field Wyman (1999:124) explains that 'a good medical index can even extend the use of a book beyond the intended audience'.

Subject's Point of View

Hazel Bell (1996:3f) refers to this category or class concerned with indexes as the 'indexed'. She reasons that people like to see their own names appear in the index of a publication in which they are mentioned. Personal names should be included in the text for valid reasons, for instance historical figures or significant personae mentioned in the text. References

should be objective, without political or any other bias. The listing of several coauthors should include all the authors, and not only the first two or three (Bell 1996:3,4).

Author's Point of View

It is often difficult to meet authors' expectations when indexing their books. Some authors list, highlight or underline the terms that they wish to have included in the index. These authors may not understand the requirements of the construction of a meaningful index.

Other authors may accept the value of the indexer's labour and will answer the questions the indexer may have regarding various issues in the text. Problems do arise in some instances when the indexer cannot communicate with the author. This type of situation will force indexers to resort to their own ideas or turn to other persons for advice.

Publisher's Point of View

The publishing editor may have specific ideas about the length or content of an index. These ideas are not always practical. Problems may also arise when the indexer is not given sufficient time to complete a properly constructed index. This may result in a rushed job which causes the indexer to become exhausted and stressed, working long hours in order to finish the index quickly.

Publishers may also be reluctant to include the introductory or explanatory paragraph at the beginning of the index. In other instances the publisher may also restrict the length of the index. The publisher may also expect the indexer to compile a 'medium level' index or an index which consists only of main headings or entries. The requirements set by the publisher are often difficult to meet and a less-than-ideal index may result.

Regulator's Point of View

Guidelines and rules are available to indexers. These serve to guide the indexer on matters concerning the format and standards set for indexes. Examples of these resources are the *Chicago manual of style*, books by individual authors, such as Judith Butcher's (1981) *Copy-editing: the Cambridge handbook*, and the British Standard's *BS 3700*: *Preparation of indexes*.

Valuable guidelines for indexing that are covered include issues such as the following.

Main Headings

'Main headings should always be nouns (qualified or not), never adjectives or verbs on their own' (British Standard).

If a word is used in both singular and plural forms in the text, only one form should be used in the index. Judith Butcher (1981) uses this example:

bishops	*or*	bishop(s)
duties		duties
incomes		incomes

But if the words have different meanings, this rule would not apply, as in the case of the words 'damage' and 'damages'. The general rule is to use plural for countable objects and singular for non-countable concepts (Calvert 1996:74).

Care should also be taken with the entries of proper names, such as saints, kings and popes, and compound names such as W Somerset Maugham, which would be entered as Maugham, W Somerset.

Sub-entries

Butcher (1981) states: '[I]f an entry contains more than six page references, or a reference spans more than nine consecutive pages, it should usually be broken down into sub-entries ... Passing mentions should not normally be indexed.'

The *Chicago manual of style* agrees with this measure, but a different standard is required by the ASI where the requirement that main headings with more than five to seven page references should be broken down into sub-entries.

Other Issues

Further requirements listed by the ASI are that sub-entries should be useful and concise, with the most important word at the beginning. Unnecessary words or phrases should be avoided.

Issues such as the application of 'broken off' or 'run-on entries', cross-references, the use of double postings or double-entering, and synonyms, need to be applied consistently and with the necessary care.

The presentation of the index can affect the usefulness of the index. The criteria for presentation and layout were discussed above.

Length of Index and Typeface Used

The ASI recommends that the length of an index should be 3 to 5 per cent of the pages in a typical nonfiction book, and perhaps 5 to 8 percent for a history or biography book and 15 to 20 per cent for reference books. Typographical devices such as bold type or italics can be used to distinguish items such as illustrations or cross-references.

Format

The requirements of the format include matters such as the size of the print, appearance and aesthetics. The punctuation and style should be consistent. Entries that run from one page to the next should be repeated, followed by the word 'continued' in parenthesis.

Apart from the rules and standards formulated as guidelines for the construction of an index, other attributes are also commendable in indexes. These attributes include user-friendliness, readability, maximum detail, and so on. An example of a description of a good index is presented by Wyman (1999):

> If you want to become educated in food science, food technology, or nutrition, here is your textbook. Its 304-p index, with a total of about 48,000 key words/terms, will

lead you to any subject you wish to search. Frankly, I'm impressed. I could not find a concept or word that was not listed.

On the other hand, a poorly constructed index is reviewed in the same article:

The only blot is a disgracefully negligent index. In a book of this size and this value, one needs a bit of help, but Viking has produced an index which is useless, just a list of page numbers. The name of Dorothy Wordsworth is followed by nearly 400 page numbers, but if you want to know anything in particular, you will have to fish around helplessly. I hope, before the paperback, Viking does one of our best biographers proud by producing an index worthy of a remarkable book (Wyman 1999:158).

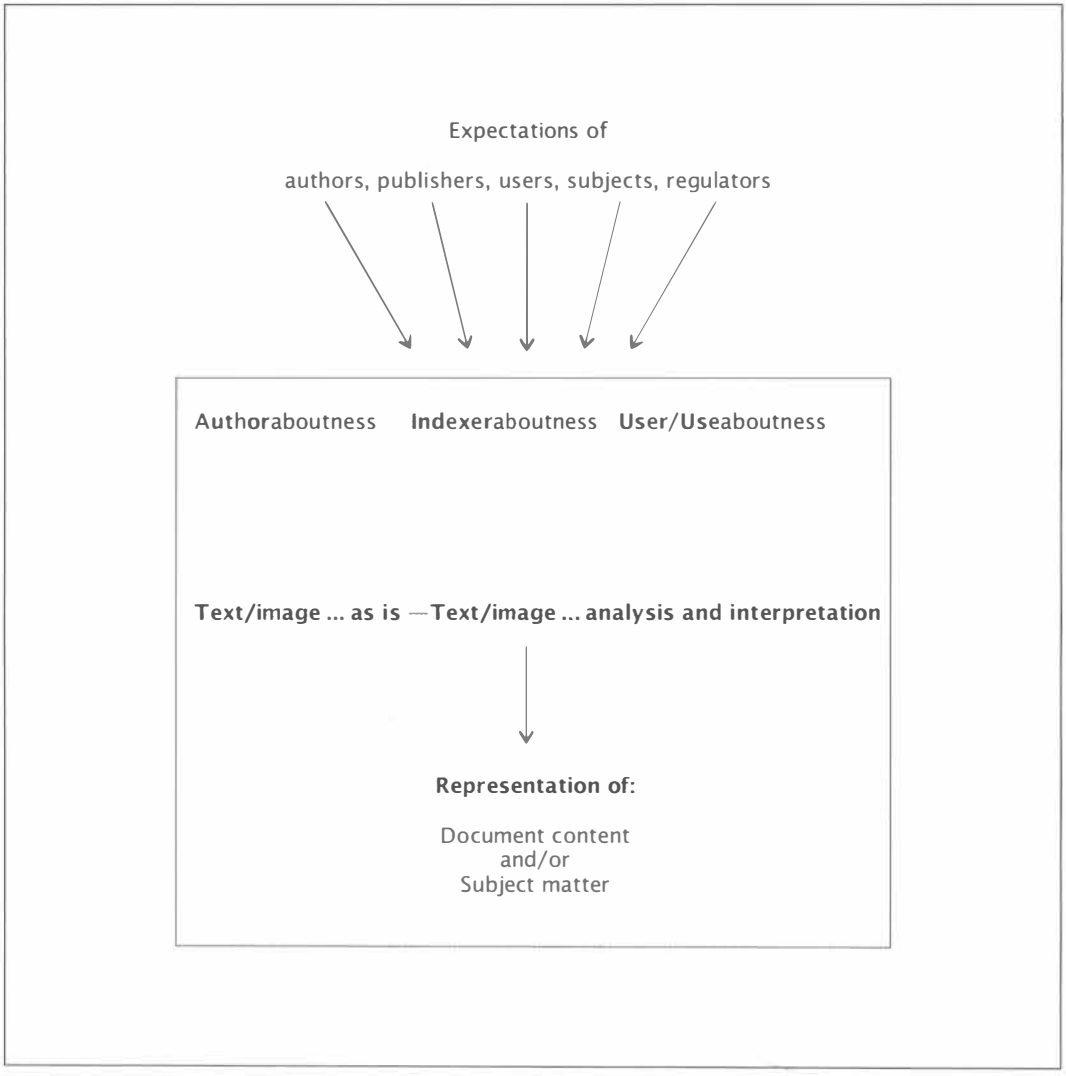


Figure 3: *The Elements Involved in and the Influencing Factors on the Indexing Process* (adapted from Ingwersen 1992:55)

Recommendations

The indexer who works systematically and meticulously through a text should be able to produce a good index. During the indexing process the indexer needs to pay attention to many criteria, expectations or challenges. The indexer may find it useful to keep a personal style manual in which to make notes and keep a record of indexing problems and decisions made. A good index includes entries to all indexable material such as the basic text, significant items in forewords, prefaces, introductions and appendices of books (Borko & Bernier 1978:41).

It may also be helpful to pay attention to some typical problems often encountered by indexers. Typical problems student indexers have are listed in a review article by Ann Hall (1997). The most significant of these include the following:

- Inconsistencies in spelling, punctuation, use of verbs, arrangement of sub-entries, and so on. The index should be organised according to a logical and consistent plan, and should also be in harmony with the subject matter and readers' expectations (Borko & Bernier 1978:41).
- Use of umbrellas, meaning entries to entries that are inclusive (such as 'fruit') and then failing to enter pages with references to individual types of fruit.
- Embedding, where entries are made to subjects where the average reader will not look, that is 'publications'.
- Double-entering should include page numbers under both entries.
- 'Woods and trees syndrome', where the index fails to reflect the structure of the publication. Hall (1997:183) explains that an index should refer not only to the details (trees) but also to the main entries (woods).

The indexer who is creative and meticulous and follows the guidelines available to indexers should succeed in compiling a worthy index. For a detailed discussion on the presentation of an index, see chapter 15.

Conclusion

The uncertainty surrounding the process of indexing hinders the formulation of rules for the extraction of concepts from the text that is being indexed. Human indexing is therefore done by the indexers based on their judgment or discretion. This impacts on the practice of the evaluation of indexes, as there are no set rules to which the index needs to comply. The indexer is also challenged by the expectations of the user, author, the subject indexed, the publisher and regulator. However, despite having no clear measures to evaluate it, the index needs to meet certain normative criteria.

The aim of teaching indexing is to equip trainees with the required skills to master the art of indexing. The mastering of these skills through training and practice would be incomplete without assessment of the indexer's competence. Evaluation of the normative criteria is essential to the teaching of indexing, to ensure that the student indexer gains the required abilities for indexing. Indexing evaluation can be beneficial in a number of ways.

It can benefit the indexer as a means of self-assessment, assist in the assessment of training courses in indexing, be applied by societies for indexers as a means to offer recognition of the competence of members of the society, help publishers to identify competent indexers for the indexing of publications, and serve to raise the standard of indexing practice and professionalism.

Indexing evaluation can be used in a constructive way by a society for indexers for the testing of its members' professional competence.

The British Society of Indexers recognises the professional competence of its members in two stages. Accredited Indexers need to have passed the Society's standard tests which indicate that they have a theoretical competence in indexing, while registered indexers need to have achieved practical competence through the Society of Indexers' index assessment procedure and have been admitted to the Register of Indexers.

The ASI does not require proof of competence for its directory and lists all available indexers in its annual directory. In South Africa, the Association of Southern African Indexers and Bibliographers (ASAIB) lists only active indexers in its directory, without any testing or evaluation. ASAIB should strive to follow the example set by the Society of Indexers and develop its capacity to evaluate and give recognition to the indexing competence of practising indexers.

BIBLIOGRAPHY

American Society of Indexers. 2001a. *Criteria for the H.W. Wilson Award.* Available: http://www.asindexing.org (accessed 17 March 2004).

American Society of Indexers. 2001b. *Indexing evaluation checklist.* Available: http://www.indexing.org/site/checklist.shtml (accessed 17 March 2004).

Bell, HK. 1996. Whom should we aim to please? *The Indexer*, 5(1):3-5.

Borko, H & Bernier, CL. 1978. *Indexing concepts and methods.* London: Academic Press.

British Standards Institution. 1988. *British standard recommendations for preparing indexes to books, periodicals and other documents.* 2nd rev ed. London: British Standards Institution. (BS 3700 1988).

Butcher, J. 1981. *Copy-editing: the Cambridge handbook.* Cambridge: Cambridge University Press:131-145.

Calvert, D. 1996. Deconstructing indexing standards. *The Indexer*, 20(2):74-77.

Chicago manual of style. 1982. 13th ed. Chicago: University of Chicago Press:511-557.

Chicago manual of style. 1993. 14th ed. Chicago: University of Chicago Press.

Collison, RL. 1969. *Indexes and indexing: guide to the indexing of books, periodicals, music, recordings, films, and other material, with a reference section and suggestions for further reading.* London: Ernest Benn.

Dixon, Y. 1996. Indexing for children. *The Indexer*, 20(1):8-15.

Farrow, J. 1995. All in the mind: concept analysis in indexing. *The Indexer*, 19(4):243-247.

Forrester, M. 1995. Indexing in hypertext environments: the role of user models. *The Indexer*, 19(4):249-256.

Hall, A. 1997. Problems, some usual (Marking Book Indexing Post Tutorials). *The Indexer*, 20(4):182-184.

Ingwersen, P. 1992. *Information retrieval interaction.* London: Taylor Graham.

Ingwersen, P. 1998. *Understanding IR: an introduction to IR*. Copenhagen: Royal School of Library and Information Science:48-60.

Kascus, MA. 1996. Indexing, in theory and practice, in *Technical services management, 1965-1990: a quarter century of change and a look to the future*, edited by LC Smith & RC Carter. New York: Haworth Press:241-278.

Lancaster, FW. 1991. *Indexing and abstracting in theory and practice*. Illinois: University of Illinois.

Manley, S & Harwood, N. 1997. Bringing it home: learning to index books by correspondence. *The Indexer*, 20(4):185-187.

Mathews, PL & Bakewell, KGB. 1997. *The Indexer*, 20(4):193.

Moys, E. 1997. Classified v. specific indexing: a re-examination in principle. *The Indexer*, 20(3):135-136.

O'Connor, BC. 1996. *Explorations in indexing and abstracting: pointing, virtue, and power*. Englewood: Libraries Unlimited.

Olason, SC. 2000. Let's get usable. *The Indexer*, 22(2):91-95.

Simkin, JE. 1997. Professionalism. *The Indexer*, 20(4):178-181.

Society of Indexers. 1991. TAB9/91. (Unpublished)

Vickery, BC. 1973. *Information systems*. London: Butterworths.

Wallis, E. 1997. Indexing as a professional activity. *The Indexer*, 20(4):187-194.

Weinberg, BH. 1994. Indexes: a chapter from *The Chicago manual of style, 14th edition*: a review. *The Indexer*, 19(2):105-109.

Wormell, I. 1992. *Understanding information*. Copenhagen: Danmarks Biblioteksskole.

Wyman, P. 1999. Medical indexes reviewed. *The Indexer*, 21(3):124-126.

5 ORGANISATION POLICY

Marlene Burger and Ina Fourie

Abstract

In this brief chapter we want to make you aware of the importance of formulating a policy for the verbal organisation of information. The focus is on indexing and abstracting.

Introduction

To ensure effective retrieval of information, it is essential to formulate a policy for the information organisation and retrieval system (IOAR). If you plan and operate an IOAR system, the onus is on you to determine the policy. On the other hand, if you make contributions to an IOAR system (e.g. a national database), you will have to understand and apply the policy laid down by others. A policy is essential; otherwise there will be chaos in the organisation and retrieval of information. The policy must be based on the purpose of the IOAR system and on approved standards.

An indexing policy lays down guidelines for how indexers should study, analyse and describe the content of entities by allocating indexing terms. An abstracting policy gives guidelines on how abstractors should study, analyse and describe the content of entities as a shortened, accurate synopsis of the original entity. The following is an example of what to consider when a policy for an IOAR system is to be formulated (note the place of indexing and abstracting):

Coverage	What type of entity will be included (e.g. books, journal articles, photos, images, maps). What subjects will be covered?
Extent	What is the expected number of entities to be covered in the IOAR system? (This will influence the choice of vocabulary.)
Users	Who will use the IOAR system (e.g. lay-people, specialists, scientists, academics, schoolchildren)?
Available technology	Will the IOAR be a manual system or a computer system, or will it be run via the Internet? Which hardware (e.g. computer equipment) and which software (e.g. database program) is available for the design of the system? To what degree can technology be used to support the organisation of information (e.g. computer-aided indexing)?
Forms of information organisation	Which form of bibliographic description will be used (e.g. AACR2, Dublin Core)? Which form(s) of verbal subject description will used (e.g. indexing, abstracting, subject headings)? For each of these, a specific policy is necessary.

Which indexing language will be used?	Will natural language or a controlled vocabulary be used? What will be used in the case of controlled vocabulary (e.g. a thesaurus, a list of subject headings)?
How will the descriptions (indexing and abstracting) be done?	Will the indexing and abstracting be done manually or will an automatic method be used?

The above forms the background against which an indexing or abstracting policy should be formulated. Such policies must take the purpose of the IOAR system into consideration, since they are going to operate as part of the system.

Indexing Policy

An indexing policy is important to the formulation of guidelines on how indexers should study, analyse and describe entities. Soergel (1985:329) contends that knowledge of the indexing policy is important to the assigning of indexing terms and for retrieving information: 'to use knowledge of indexing criteria for writing query formulations that exploit all the possibilities offered by the indexing and also to compensate, to the extent possible, for any weaknesses in indexing'. An indexing policy pays attention to the following:

- the type of information that is important and should be highlighted (e.g. in a database about nuclear power, any reference to techniques that uses radio-isotopes is very important, even it there is only limited information about it)
- the type of information that should not be covered (e.g. historical information or information that is generally known)
- the specificity of indexing (i.e. assigning specific or general indexing terms)
- the exhaustivity of indexing (i.e. the depth of indexing)
- the point of departure for indexing: user needs (thus the intellectual content of entities) (see chapter 3 for more information)

An indexing policy is important to setting up a standard of indexing and to ensuring that indexing terms are assigned consistently. The following standards can be used as a basis for a policy:

BS ISO 999: 1996. *Information and documentation: guidelines for the content, organization and presentation of indexes.* London: British Standards Institution.

BS 6529: 1964. *Recommendations for examining documents, determining their subjects and selecting indexing terms.* 1984. London: British Standards Institution.

The specificity of indexing terms and the exhaustivity of indexing are two important issues for an indexing policy when it comes to users' needs. In the following two sections these concepts are discussed in more detail.

Specific versus General Terms

Specificity is the generic level on which the concepts assigned to the entity are expressed. It refers to the degree to which the indexing terms assigned correspond with the precise meaning of the concepts described by the entity. If the indexing terms correspond with the concepts described, one speaks of *specific indexing*. If more general terms are used to describe the concepts, they are referred to as *general indexing terms*.

If very specific indexing terms are used, users are in a position to look for information under specific terms. Users do not then have to look through other information concerned with more general terms. If they require information about African elephants, they do not need to go through all the entities containing information about elephants in order to select specifically those containing information about African elephants.

If natural language is used for indexing, the indexing terms can be as specific as the terms found in the entity. If a controlled vocabulary (e.g. a thesaurus) is used, the specificity of the indexing terms depends on the specificity of the thesaurus vocabulary. The following figure is an example of two thesaurus entries. The one contains very specific indexing terms, while the other contains general indexing terms. The terms indicated here thus determine the terms which will be used for indexing and for retrieving information.

Thesaurus A with general entries	**Thesaurus B with specific entries**
LIBRARIES	**LIBRARIES**
RT Archives	NT Academic libraries
Community centres	Public libraries
Museums	Special libraries
	RT Archives
	Community centres
	Museums

Exhaustivity of Indexing

Taylor (2004:250) puts it like this: 'Exhaustivity is the number of concepts that will be considered in the conceptual framework of the system. The number of concepts any given agency's cataloguers/indexers will assign is often guided by local policy.' Lancaster (1998) describes it as the extent to which the breadth of subject matter discussed in a particular document is covered in a representation of that document. It will correspond roughly to the number of index terms assigned or some other measure of access points provided. Cleveland and Cleveland (1990:111) refer to exhaustivity as the depth of indexing.

The exhaustivity of indexing is thus related to the extent to which indexing terms are assigned to the features covered in an entity. The issue is whether indexing terms are only assigned to a few main concepts covered in the entity (low level of exhaustivity), or whether indexing terms are also assigned to features which the entity deals with in some depth (e.g. a few paragraphs) (high level of exhaustivity).

The degree of exhaustivity of indexing is not necessarily related to the number of indexing terms assigned. For example, a lengthy document dealing with the 'advantages and disadvantages of the use of the *Dewey decimal classification* in special libraries' may

only be assigned two indexing terms: *Dewey decimal classification* and *Special libraries*. The indexing may still be exhaustive. Another document may, for example, deal with the 'advantages and disadvantages of the use of various classification systems in special libraries', and for this several more indexing terms will have to be assigned to index the document exhaustively. For example Dewey decimal classification, Library of Congress classification, Universal decimal classification and Special libraries.

The level of exhaustivity at which indexing terms are assigned is determined by the indexing policy. While the level of exhaustivity is not related to the characteristics of the indexing language, it may be related to the type of entity indexed (internal reports, for example, will be indexed more exhaustively), as well as the subject of the entity (as indicated in the example of classification systems).

The exhaustivity of indexing also influences the retrieval of entities:

- A high level of exhaustivity leads to a high level of retrieval. All possible entities containing information (even if it is only a little) are retrieved.
- A low level of exhaustivity leads to a high level of precision. All entities retrieved should contain a reasonable amount of information about the subject.

Exhaustive indexing is particularly useful for users who require an overview of all the entities on a subject, since several entities will also be retrieved that may only contain a few paragraphs on the subject. It is less useful for users seeking specific facts or information.

The following (adapted from Soergel 1985:328) is an example of instructions to indexers (contained in the indexing policy) about intended exhaustivity:

Low exhaustivity (few descriptors)
- Use an indexing term only for entities that are highly relevant for a search about an index term.
- Only index the key subjects of the entity.
- Index, for example, only the first ingredient of food.

High exhaustivity (more descriptors)
- Use an indexing term for entities that are only marginally important.
- Index less important (minor) subjects covered by the entity.
- Index, for example, all the ingredients of food.

Abstracting Policy

An abstracting policy gives guidelines to abstractors on how they should study, analyse and describe the content of entities, so as to give an accurate, abbreviated version of the original entity. Such a policy is necessary to maintain the standard and to promote consistency among abstractors. It should take the following into consideration:

- unique nature of the user group
- purpose of the abstracts
- nature of the original entities
- characteristics of abstracts

- guidelines laid down by standards of abstracting
- purpose of the IOAR system

An abstracting policy should meet the following requirements:

- An introduction to the abstracting service, database or other IOAR systems should be included, as well as the purpose of the abstracts and how they will be used. (It is important to remember that abstracts can serve as search points when retrieving information.)
- Criteria to use in selecting entities for abstracting.
- Guidelines on how to analyse an entity.
- Guidelines about the type of abstract (e.g. indicative, informative, critical). They can also determine that a combination of indicative and informative abstracts should be written for reports, for example.
- Guidelines about the type of information that is regarded as important (e.g. purpose of a study, the steps followed, the methods used and the findings).
- Guidelines about information that may not be included (e.g. references to specific table numbers or information not covered in the entity).
- Instructions about the physical presentation and layout of the abstract. These can include instructions about the style and length, that the abstract should be written in one paragraph, or that it can be presented as a structured abstract under different headings.
- The language that the abstract should use and the handling of abstracts in foreign languages.
- Guidelines for abstracting procedures (i.e. how to prepare the abstract).
- Guidelines for special approaches to be followed for different types of entity.
- Guidelines about the form of the bibliographic description and the information that must be included.
- Lists of standard abbreviations that may be used.
- Examples of abstracts for different topics.
- Guidelines for the editing of abstracts.
- Reference(s) to preferred standards to be used, such as the American National Standards Institute's (*American national standard for writing abstracts.*) 1979. New York: ANSI Z39.14:1979).

In order to understand the context of information organisation policies, you must also consult chapters in this book dealing specific issues, such as those in Part Two.

Conclusion

Indexing and abstracting policies are essential to give guidelines to indexers and abstractors. They improve the quality of the work and promote consistency. Keep in mind though that policies are seldom drawn up by individuals – it is mostly a team effort. Policies should also be revised or adapted regularly to incorporate new technologies, changes in the IOAR overall policy, and so on.

BIBLIOGRAPHY

Cleveland, DB & Cleveland, AD. 1990. *Introduction to indexing and abstracting.* 2nd ed. Englewood, Colo: Libraries Unlimited.

Cleveland, DB & Cleveland, AD. 2001. *Introduction to indexing and abstracting.* 3rd ed. Englewood, Colo: Libraries Unlimited.

Fidel, R. 1994. User-centered indexing. *Journal of the American Society for Information Science*, 45(8):572-575.

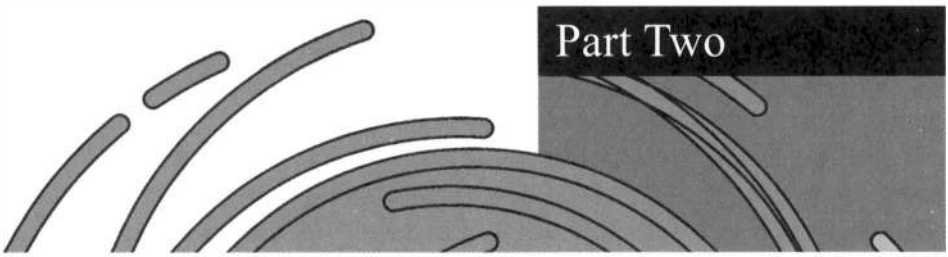
Lancaster, FW. 1998. *Indexing and abstracting in theory and practice.* 2nd ed. London: Library Association Publishing.

Lancaster, FW. 2003. *Indexing and abstracting in theory and practice.* 3rd ed. Champaign: University of Illinois, Graduate School of Library and Information Science.

Rowley, J & Farrow, J. 2000. *Organizing knowledge: an introduction to managing access to information.* 3rd ed. Aldershot, UK: Ashgate.

Soergel, D. 1985. *Organizing information: principles of data base and information retrieval systems.* San Diego: Academic Press.

Taylor, AG. 2004. *The organization of information.* 2nd ed. Westport, Conn: Libraries Unlimited.



GENERAL APPLICATIONS

6 VERBAL SUBJECT DESCRIPTION

Ina Fourie and Marlene Burger

Abstract

The intricacies of verbal subject description are discussed. These include a selection of types of verbal subject descriptions, the difference between precoordinate and postcoordinate indexing, decisions to take in verbal subject description and the problems associated with the use of vocabulary. Reference is made to the various verbal methods discussed and analysed in this book.

Introduction

Information retrieval systems (IRS) such as databases, library catalogues, bibliographies, indexing and abstracting journals, portals and book indexes are created to help people who need information to find the desired solution to the problem. According to Ingwersen (1992:228) an IRS 'is constituted by interactive processes between its *system objects, system setting,* and the environment, capable of searching and finding *information* of potential value to an actual searcher of information'. To this can be added Harter's (1986:245) argument that an IRS should save time and effort. It should retrieve the records that are required (relevant) and withhold records that are not required (not relevant). He describes an IRS as 'a device interposed between an *end-user* of an information collection and the collection itself ... [whose purpose] is to capture wanted items and filter out unwanted items from the information collection'. To ensure that an IRS fulfils its purpose, records are created (also called surrogates, document representations or entity representations). In these records the documents or other entities (such as people, museum objects, websites) are physically described, as well as according to their intellectual content (what they are about, or aboutness). To describe the intellectual content one can use the following:

- notations where numbers and other symbols are used
- verbal subject descriptions where words are used

For the latter, vocabulary is used which can be selected from an indexing language that can be based on natural language or it can be prescribed by a controlled vocabulary such as a thesaurus. Abstracts and subject headings are also verbal representations of the subject content of entities.

When indexing periodical articles, each article will be represented by a bibliographic record (physical description). For example:

Fourie, I. 2001. The Use of CAI for distance teaching in the formulation of search strategies. *Library Trends*, 50(1):110-127. (Reprinted from *Mousaion*).

The physical description is prepared according to guidelines and standards such as the *Chicago manual of style* or the *Anglo-American cataloguing rules.* There are also guidelines for the description of the intellectual content (e.g. guidelines, as well as standards for indexing, abstracting and the allocation of subject headings) (Cleveland & Cleveland 2001; Lancaster 2003).

The quality of the physical as well as intellectual descriptions affects the efficiency and effectiveness of the IRS, and therefore the preparation of descriptions of the intellectual content of entities is often considered to be one of the most difficult and complex tasks. Blair (1990:vii-viii) puts it as follows: 'The central task of Information Retrieval research is to understand how documents should be represented for effective retrieval. This is primarily a problem of language and meaning. Any theory of document representation, and therefore Information Retrieval, must be based on clear theory of language and meaning.' According to Hjørland and Nielsen (2001:250): '[A]ccess points determine in a rather firm way the objective possibilities that are provided for the talented user (or for any formalized, algorithmic, or automatic procedure). Therefore, it is essential in information science (IS) to develop knowledge about what kinds of subject data exist as well as the strengths, weaknesses, and relative contributions of each kind.'

The description of the intellectual content by means of words is called *verbal subject description.* Verbal subject description is preceded by a process known as conceptualisation, discussed in chapter 3. Verbal subject description can be based on human, intellectual input or automated or semi-automated methods. More detail on automatic indexing and abstracting can be found in the relevant chapters. As words form the basis of this process, one is faced with many linguistic problems such as meaning, synonyms, bias, phrases, compound terms and lay versus scholarly use.

Notations versus Verbal Subject Descriptions

Notations such as the *Dewey decimal classification* (DDC) numbers and verbal subject descriptions, like abstracts, descriptors, subject headings, identifiers, title indexing and PRECIS strings, have both advantages and disadvantages. However, it is important to understand that neither classification notations nor verbal methods of subject analysis can encompass all aspects of subjects in the majority of entities. Even if as many notations, subject headings, indexing terms or descriptors are used as required, these would still represent only the principal subjects of the entity in question. (Macroanalyis is used when an entity is analysed and described in its entirety, or when only the main subject matter is singled out; microanalysis is used when parts of an entity are analysed (e.g. a chapter in a book or articles in a journal) and the subject matter is examined in more detail or in greater depth, i.e. depth analysis). The following examples of an entity on 'computer-assisted instruction in distance teaching in search skills' illustrate the difference.

Examples

The following is an example of a notation:

DDC	374.26 (use of computers in adult education; electronic distance teaching)

Other classification systems such as the *Library of Congress classification* (LCC) and the *Universal decimal classification* (UDC) will have different notations to express the same subject.

Examples of verbal subject description include the following:

Indicative abstract	Discusses the use of computer-assisted instruction (CAI) in the formulation of search strategies, and its potential for adult, distance students. The program is used with students from the University of South Africa (Unisa), a distance teaching university. Two versions of the program (1992, 1998) are analysed according to the design teams' experiences and student feedback (collected through questionnaires). The actual design is discussed according to instructional design principles. Selected student feedback is summarised.
Informative abstract	Discusses the use of computer-assisted instruction (CAI) in the formulation of search strategies. The program is used with students from the University of South Africa (Unisa), a distance teaching university. The importance distance teaching holds for life-long learning and programs for adult learners, as well as the potential of CAI for distance teaching is considered. The program was designed by the Department of Information Science. The first version was introduced in 1992, and the second in 1998. They were based on literature reports on the use of CAI in Library and Information Science, as well as principles of instructional design. The latter included determining the need for the CAI, a situation analysis, analysis of the learning content required for the formulation of search strategies, and an analysis of the learners and the availability of technology and support. Objectives and outcomes were set, and methods selected for performance assessment. These include a pretest on two levels of difficulty: Refresher and Challenge. Aspects considered for the actual design, development, implementation and distribution of the program are briefly covered. As part of the formative assessment, feedback was collected on students' experiences, which were mostly very positive. They pointed out sections requiring more examples or further explanation, and the need for online guidelines as part of the tutorial.
Indexing terms, keywords (natural language)	Adult students Computer-assisted instruction Distance students Distance teaching Formative evaluation Instructional design Search strategy formulation Summative evaluation

Descriptors (selected from a thesaurus, e.g. *Thesaurus of ERIC descriptors*)	ADULT STUDENTS COMPUTER-ASSISTED INSTRUCTION DISTANCE EDUCATION FORMATIVE EVALUATION INFORMATION SEEKING INSTRUCTIONAL DESIGN SEARCH STRATEGIES SUMMATIVE EVALUATION
Identifiers (these are terms that are important enough to select as indexing terms, but do not appear in the controlled vocabulary such as the *Thesaurus of ERIC descriptors)*; they are used in systems based on controlled vocabulary.	Performance assessment South Africa University of South Africa Etc.
Subject headings (there are different lists of subject headings available e.g. *Sears list of subject headings, Library of Congress subject headings [LCSH], MESH*)	Adult education - South Africa Computer-assisted instruction Distance education - South Africa Information science - Study and teaching Etc.

The following can also be considered as examples of verbal subject descriptions: broad categories, for example as used by some search engines and directories, titles, PRECIS strings and notes.

Brief Comparison

The methods of subject expression illustrated above are compared briefly so as to indicate relationships and differences. Please note that all of these flow from the initial *conceptual analysis* phase necessary for effective subject description.

Classification Notations

Instead of words, subject matter is expressed by numbers – notations are not language-bound, that is the meaning of a notation is the same in all languages. The function of the classification notation and the terms resulting from verbal subject analysis is different, even though the words describing the notation are the same as those of the subject heading, and so on. For example, the subject heading verbalises specifically a major subject in an entity, while the notation indicates where in the arrangement of its system that subject is to be placed to determine its relationship (usually hierarchical) to other subjects within the discipline. A disadvantage of using notations only is that the user does not necessarily have

knowledge of the classification system and therefore cannot interpret the notations. An index to the notations has then to be compiled as a 'key' to these.

Abstracting

Abstracts give short, concise, accurate descriptions of the intellectual contents of entities so as to promote awareness of entities and information retrieval. Abstracts are often approached through an index, especially if the abstracts have a classified arrangement. They are not meant to replace entities, but serve as surrogates to aid decision making on relevance for a certain information need. Abstracting requires certain language skills and familiarity with subject content. Disadvantages are that they may be affected by errors, policy, omissions and abstractor bias, or that they may be so poorly written that they are difficult to use. It may also be that not all users are equally proficient at using abstracts and other bibliographical tools, so that the user imposes self-limitation on the tools (Cleveland & Cleveland 2001:60).

Indexing

Indexing entails the assigning of indexing terms (also known as descriptors) by using the natural language of the entity or the subject discipline, or by using words from a controlled vocabulary such as a thesaurus. The terms can be combined by Boolean operators or in other ways during the retrieval process to locate the desired information. Citation indexing refers to the situation where the author quoted in an entity (e.g. bibliography, list of cited sources) is used as the point of departure to indicate relationships with other entities. The rationale is that the entities quoted must be on the same subject(s), and that they must be important to the specific subject(s). Book indexes are normally based on the text of a particular entity (i.e. the indexer must use the words appearing in the text); book indexers must therefore have sufficient subject knowledge to add terms that may be useful to the reader. Furthermore, such indexes use natural language, which has its own problems. Terminological control is essential to successful retrieval. Natural language, because of the variety in subject terminology, may cause retrieval problems. To overcome this, some kind of control is necessary: either at the indexing input stage or at the search stage. Book indexes are usually prepared by one person; periodical indexes are compiled by a number of people over a number of years, with shifts in subject emphasis and indexing objectives. Each issue of a periodical may deal with unrelated topics by several different authors, written in different styles, aimed at different users. The periodical index must bring order out of these divergences.

Thesauri

These are controlled vocabularies. They are arranged alphabetically or hierarchically (i.e. formally structured) and can be used for a computerised database or a printed index and are available in printed or electronic format (online thesauri). Thesauri indicate a variety

of indexing terms occurring to the indexer/user. Some of these terms are used as preferred terms in the thesaurus, while others are non-preferred or entry terms. Classification systems are generally hierarchical with secondary alphabetical indexes, while thesauri are generally alphabetical with a hierarchical structure reinforced by cross-references. The relationships between terms in a thesaurus are more specific. In thesauri, descriptors are often dependent on these terms and are intended to be combined with other terms (to express more specific concepts), whereas in a classification system or a subject heading list, the terms can stand alone. In other words, we have on the one hand pre-coordinated systems (e.g. classification notations and subject headings) and on the other hand a post-coordinated language (e.g. descriptors and combinations). In a sense, a thesaurus is a specific language, with words, relationships of words and its own grammatical rules for usage. It is a controlled, collected subset of natural language. A thesaurus is not a dictionary – a dictionary records the standard usage of words and terms, while a thesaurus is the result of the understanding of proper usage. A disadvantage is that not all terms an indexer may need are represented in a specific thesaurus.

Subject Headings

Assigning subject headings by using a standard list is the most common practice. Such lists appear in printed as well as electronic format and are used as standards for the compilation of subject headings. A subject heading list is an alphabetical list of words or phrases that can be used to describe the subject matter of entities. Some lists are general, covering most topics (such as in a general classification system), while others were developed for specific disciplines or subjects. In practice, subject heading lists may be expanded and/or adapted to meet the specific needs of an information agency. Subject headings are used as entry points (access points) in bibliographic databases, catalogues, indexing journals, bibliographies, the Internet and the like. Sometimes subject headings are used in conjunction with indexing terms. It is therefore necessary to understand the difference between thesauri, indexes and subject headings (see the examples above). As with other methods of verbalisation, a disadvantage is that the indexer selects the subject headings based on the text and his or her knowledge of the subject, but these may not concur with the perception of the searcher. *See* and *see also* references are used to guide the user to the 'correct' subject heading(s).

Precoordinate and Postcoordinate Indexing

To understand verbal subject description it is important to distinguish between precoordinate and postcoordinate indexing. According to Harter (1986:247) a precoordinate indexing system is based on a controlled vocabulary in which specific, complex concepts involving a combination of two or more concepts are used as an indexing term in the vocabulary. Coordination takes place at the time of indexing (i.e. at the time of creating the entity representation). 'Database producers' for example combine the concepts of 'databases' and 'producers', or 'computer-assisted instruction' combines 'computer-assisted' and 'instruction'. Subject headings and classification systems are examples of precoordinate indexing systems. In the previous examples used above, DDC, *Eric thesaurus* and

Sears list of subject headings were used for the article on computer-assisted instruction. The relationship between the concepts is thus indicated at the time when the entity representation is created. The relationship is determined by the person creating the entity representation. When retrieving information the same order should be used (e.g. 374.26 to retrieve information about computer-assisted instruction in adult education, or Computer-assisted instruction). Only one order and combination of terms/concepts is possible – in classification this is known as citation order.

Postcoordinate indexing systems can be based on natural language or controlled vocabulary terms/concepts. Coordination of terms/concepts takes place at the time of retrieval, that is the individual concepts such as 'database producers' and 'bibliographic databases' and 'South Africa' can be combined at the time when information is needed. This can be done by using Boolean operators: database producers AND bibliographic databases AND South Africa. Pre-coordinated terms such as 'database producers' can also be used in a postcoordinate system. The relationship between terms/concepts is not indicated at the time the record is created, but rather at the time of information retrieval. Any number of combinations can be used during information retrieval. For example: **database producers, database producers AND full-text databases**, or **database producers AND full-text databases AND South Africa**. It is also important to note that the relationship is indicated by the information seeker. Postcoordinate indexing was made possible by the early uniterm concept of Mortimar Taube, and especially the development of electronic databases and Boolean logic (Chu 2003). The following is but a brief description of the Boolean operators. There are many publications from which more information can be obtained – the book by Taylor (see bibliography) is a good starting point.

Boolean Operators

There are three Boolean operators: AND, OR, NOT. These operators are used to link search terms. Each of the operators has a specific effect on the search results, and one should therefore make sure they are well understood, because using the wrong operator can make a great difference to search results.

'AND' Operator

If two or more search terms are linked with the AND operator, it means that all the terms must appear in every record that is retrieved. For example: **Libraries AND South Africa.** Records in which only one of the search terms occurs, for example, will not be retrieved. Any number of search terms can be linked with the AND operator. For example: **Libraries AND South Africa AND CD-ROMs.** The more terms linked in this way however, the fewer records will be retrieved since the search becomes very specific – it may even happen that no records are retrieved.

'OR' Operator

If two or more concepts are linked with the OR operator, it means that records will be retrieved if any one, or even more, of the search terms occurs in the record. The more search terms linked with the OR operator, the more records can be retrieved. For example:

Libraries OR South Africa. Any number of terms can be linked with the OR operator, but note that as search terms are added, more records will be retrieved each time.

'NOT' Operator

If two search terms are linked with the NOT operator, the search terms that are preceded by the NOT operator will be excluded. Thus no records in which the search term occurs will be retrieved. This operator should be used very carefully since it can cause useful information to be lost if the search term that has to be excluded occurs together with substantial information. For example: **Libraries NOT South Africa** (no records on South Africa will be retrieved).

Truncation symbols such as ? or * are used to indicate that only the word stem is to be sought. All words beginning with the word stem are then retrieved, for example **manag?** retrieves **management, manager, managers, managing,** and so on. Some information systems also permit wild-card truncation, for example **col?r** retrieves **color, colour**. If more than one type of Boolean operator is used to connect search terms, *parentheses* are necessary to indicate how and in what sequence terms should be combined. If parentheses are not used, the default sequence is used. It is often the following: NOT, AND, OR. For example **dogs OR cats AND breeding** will be used in retrieval as **dogs OR (cats AND breeding)** and **(dogs OR cats) AND breeding** will be used in retrieval as **(dogs OR cats) AND breeding**. *Proximity operators* are used to indicate the position and sequence of words. There are various symbols for proximity operators, and information systems differ in the way in which they indicate these operators. Proximity operators are often used in indexes and some search engines, but less often in library catalogues. There are usually two proximity operators, for example an operator that specifies that words must occur in precisely the same order but that, up to a maximum of the number of words that are specified, may appear between the words – so for **academic (2w) library**, records on academic libraries, academic medical libraries and academic law libraries will be retrieved; the second operator specifies that the sequence of the words is not important, and may, up to a maximum of the number of words that are specified, appear between the words – so for **database (2n) design**, records on database design, design of databases, design of bibliographic databases will be retrieved.

Subject Headings

The various methods of indexing, thesaurus construction, abstracting, and so on, are the main focus of this book, but the assigning or allocation of subject headings (also a form of verbal subject analysis) is not covered in a separate chapter. Subject headings are used widely to access and retrieve information from bibliographic resources; therefore a brief section on subject headings is included in this chapter. Always keep in mind that when using a list of subject headings, the *principles* of the list must first be thoroughly studied .

Standard subject heading lists do not only contain subject headings, but also provide guidelines and examples for allocating subject headings. A subject heading can be described as a word, composite term or group of words (phrases) that represent a subject under which all entities dealing with the same subject can be arranged together in, for example,

a bibliography, library catalogue, or index. The physical appearance of subject headings differs from that of thesaurus descriptors and indexing terms because they sometimes contain subdivisions, qualifiers, dashes, parentheses and so on. Examples of subject headings:

Accidents – Prevention
Belief and doubt
Bestsellers (Books)
Bridge (Game)
Cataloguing – Rock drawings, paintings and engravings
English as a second language
Indexing
International Standard Book Numbers
Pretoria (Gauteng) – Pictorial works
Sewing – Dictionaries
Shakespeare, William, 1654-1616 – Quotations
South Africa – Population – Statistics
World War, 1939-1945

Purpose of Subject Headings

The primary purpose of subject headings is to identify entities on a given subject, but they also have other purposes. For example:

- To identify entities that deal with related subjects. For example **Children's literature** will also be linked to **Children's libraries** and **Libraries and schools**.
- To bring together entities dealing with the same subjects regardless of whether different terminologies are used to describe subjects. **Library technicians** are also used for entities that deal with library assistants, library clerks and paraprofessionals.
- To show the relationships between subjects or disciplines. **Botany** is indicated as related to **Biology, Natural history, Plants** and **Science**.
- To give access to any subject at any level of analysis (from general to specific). For example **Academic libraries** and **Church libraries** are considered to be specific terms, while **Libraries** are considered to be a more general term, and most general terms are **Books, Documentation** and **Books and reading**.
- To give access via a vocabulary that a large group of users have in common, regardless of whether they are specialist or lay users. This purpose is achieved by cross-references from terms not used to terms used (preferred terms), for example 'Flora *see* **Botany**'.

The principles underlying the allocation of subject headings are similar to other verbal methods of subject description: namely the user as focus, the collocation of entities on the same subject, the use of the most representative terms, and specificity.

There are two main types of subject heading. One is *true subject headings*, which indicate the subject matter of entities (topical subject headings, e.g. **Ball games**); name subject headings, e.g. **Shakespeare, William, 1564-1616**; title subject headings, e.g. **Bible.**

N.T.). The other is *form subject headings,* which do not indicate the subject matter of entities, but rather the form in which they are presented (publication formats, e.g. **Atlases, Bibliographies**; literary genres, e.g. **Poetry, Fiction**).

General Guidelines

The following points of departure need to be considered:

- Number of subject headings. Often one subject heading is sufficient, but sometimes it is necessary to allocate more than one subject heading to properly represent the subject of an entity. If the subject matter is not properly reflected by the subject heading(s), valuable information can be lost. In practice there is often a limit on the number of subject headings per entity. With online systems there is no need to impose any limitations.
- General versus specific subject headings. Do not allocate both a general and a specific subject heading if one subject heading encompasses or includes the specific one. You would not, for example, allocate both **Libraries** and **Church libraries** to the same entity.
- Multiple concepts. More than one subject heading is usually necessary for entities dealing with a number of subjects that cannot be represented by a single subject heading. For example, for an entity dealing with Zulu and Afrikaans literature, one would assign **Zulu literature** and **Afrikaans literature** since they are two different types of literature. For an entity dealing with Greek and Latin literature a subject heading inclusive of both, namely **Classical literature**, may be enough. When an entity deals with four or more concepts which are all part of a broader concept, only one subject heading (the broader concept) is needed. For example, for an entity dealing with the history of Argentina, Brazil, Chile and Ecuador, the subject heading will be **South America – History**.
- Multiple elements. An entity with multiple elements is one that covers a central subject from different perspectives and a multifaceted entity covers different elements like form, place and time. If it is difficult to find subject headings expressing these elements and facets, it is often necessary to construct a new subject heading (according to the prescriptions of the list being used). Always consider users' needs in such instances – these needs will decide whether all the elements identified in the concepts should be represented.

The general guidelines to be followed are useful when one formulates a policy for allocating subject headings. A policy is necessary to ensure that different indexers and subject cataloguers allocate subject headings consistently. When using a specific subject headings list, always check how the general guidelines provided should be interpreted and applied. Note the following:

- Only subject headings that reflect general usage or technical terminology should be used.
- Allocate as many subject headings as necessary to describe the subject matter of an entity (taking into account restrictions and system limitations).
- The plural form of nouns is used in countable objects or entities.
- Subject headings are compiled in the language of the bibliographic database or catalogue.

- Different symbols (punctuation) are used in a subject headings list for clarity: they should be used consistently.
- Three types of references are possible: *see* references, *see also* references, and general references.
- Qualifiers and subdivisions are sometimes used.
- There are different types of subject headings.
- Words, phrases and composite terms can be qualified.
- Note special guidelines for personal, place and geographical names.
- Make sure which subject headings may be added by the indexer or cataloguer, for example proper names (e.g. persons, places, battles) and general names (e.g. animals, fruit, diseases).

Necessary Decisions

As indicated earlier, the quality of the verbal descriptions influences the quality of the information retrieval. The following points need to be considered:

- Type of verbal subject description (e.g. abstract, indexing terms, subject headings). Remember this description offers the words that can be searched during information retrieval, and also allows one to decide on the relevance of an entity.
- Length of the description. This will influence the number of words that are available for information retrieval. Longer descriptions such as informative abstracts can facilitate information retrieval. They are, however, more time-consuming to prepare. Lancaster (2003) offers an extensive discussion on the impact of length of description. Hjørland and Nielsen (2001:284-285) also support the idea that a combination of different types of verbal descriptions will contribute to more effective and efficient information retrieval, and will help information seekers to decide on the relevance of entities.
- Specificity. How specific will the words be? In other words, will the words in the representation be as specific as the words used in the text? (This mostly applies to indexing terms, descriptors and identifiers.)
- Exhaustivity. How exhaustive is the coverage of the intellectual content? Exhaustivity is sometimes reflected in the number of indexing terms assigned: the use of more indexing terms is often associated with more exhaustive indexing, although this is not necessarily always the case. Exhaustivity can also be reflected in the length of an abstract and the amount of detail provided (e.g. the difference between indicative and informative abstracts).
- Vocabulary used (e.g. natural language or controlled vocabulary). If controlled vocabulary is used a suitable thesaurus should be selected or constructed according to user needs.

Language Issues

Language is always a key component in information retrieval and especially in verbal subject description. There are, however, many problems with the use of language. Note that language is used to

- represent the content of documents or other entities
- represent the information problems of the users (information seekers)
- instruct the computer to carry out the search and retrieval functions

Languages are Classified by

- vocabulary: strings of continuous characters that form words (e.g. student, student performance, AIS303)
- syntax: rules for combining words to form valid sentences or phrases (philosophy of education; education of philosophy; of education philosophy [not valid])
- logical structure (or taxonomy): describes the nature of the relationships between elements
- domain: can be either a real world or a theoretical world (e.g. discipline such as History or Computer Science, or a work environment like a group of engineers)

Natural language and controlled vocabulary are both indexing languages that are used in verbal subject description. Natural language refers to the words of the author that can be found in the title and full text of an entity, the words used every day in the subject field, or words that an indexer can select from his/her own vocabulary on the subject that is discussed in the entity. Controlled vocabulary refers to terms that are selected from a list of possible indexing terms that has been compiled beforehand (e.g. a thesaurus or a list of subject headings). In the case of indexing terms, we refer to terms selected from a thesaurus as descriptors (e.g. selected from *Thesaurus for ERIC descriptors*).

Both natural language and controlled vocabulary have advantages and disadvantages. The following section is based on the work of Aitchison, Gilchrist and Bawden (2000), Chu (2003), Harter (1986), Lancaster (2003) and Soergel (1985).

Natural Language

The advantages of natural language pertain particularly to the organisation of information (representations of entities) and the disadvantages to retrieval (search terms and search queries).

The following are examples of the advantages of natural language:

- Very specific indexing terms can be assigned. The terms can be assigned as specifically as they are used by the author. If the author of an entity refers, for example, to Persian cats, it can be assigned as an indexing term; a more general term such as 'cats' need not be used.
- Natural language represents reality. The terms used in the everyday world (reality) can also be used as indexing terms. (This is in comparison to controlled vocabulary that is often described as an artificial language.)
- New terminology can easily be added to the vocabulary because it can be chosen from the content of an entity or can even be assigned. It is not necessary to wait until terms have been added to lists of controlled vocabulary.

- It is easier to adapt natural language according to changes taking place with regard to the terminology of the language.
- Natural language makes it possible to keep the differences in meanings of terms (e.g. synonyms and near-synonym terms such as 'diet' and 'nutrition' or 'inflation' and 'price increases'). This is in contrast to controlled vocabulary where a maximum of one or two terms can be used.
- The translation process (thus the process of subject description) requires less time and effort and less utilisation of personnel. It is not necessary to first consult a list of controlled vocabulary or, if a list is not available, to find alternatives. Concepts are translated directly by using the words or phrases of the author, or the words or phrases used by the indexer. It requires less mental effort because it is not necessary to think of alternative words if a specific term does not appear in a list of controlled vocabulary.
- The use of natural language requires no maintenance, as is the case with a list of controlled vocabulary.
- Using natural language to retrieve information makes it very easy, because people can use the terms that they thought of themselves.
- It is not necessary for people who wish to retrieve information to receive special training on how to use a list of controlled vocabulary such as a thesaurus, subject heading list or a classification system.

Most of the advantages mentioned above are concerned with the organisation (thus indexing) of entities. The disadvantages pertaining to the use of natural language are concerned with the retrieval of information:

- The use of natural language requires more work than when controlled vocabulary is used. Although a more specific search for information can take place (remember any term can be used; it does not have to be in a controlled list), those who need the information must think of all the possible terms themselves, instead of using the indexing terms which were assigned during organisation. These may be referred to as search terms. A number of disadvantages include
  - variations in spelling
  - plurals (singular and plural can be used for search terms)
  - different word forms (nouns, verbs or adjectives can be used)
  - synonyms or near-synonym terms that may help to retrieve relevant information
  - broader terms (or concepts), for example the use of 'cats' if a search under 'Persian cats' (a more specific or narrower term) does not deliver adequate information
- To cover all these possibilities, users often have to use dictionaries, reference works, thesauri, classification systems and other lists of controlled vocabulary to ensure that the information required is retrieved.
- There are terms that are difficult to express using indexing terms in natural language, because they are actually descriptions rather than terms. Consider, for example, how you would search for information in a database about cancer (CANCERNET) if you were interested in survival for more than five years. In such a case it would be far easier if an appropriate term in a list of controlled vocabulary could be used.

- In multi-language databases the use of natural language can give rise to problems.
- In some instances, and particularly in the case of full-text databases, the use of natural language can result in too much information being retrieved. (Full-text databases contain the full text of an entity, e.g. journal articles, newspapers and encyclopaedias).
- Information about the same subject is often found scattered in an alphabetical list of terms, because different terms are used. For example 'indexing', 'subject indexing' and 'subject description'.
- If a subject is mentioned implicitly rather than explicitly in a text, it will be more difficult to ascertain using derived indexing (natural language).

Controlled Vocabulary

Controlled vocabulary has advantages and disadvantages. The advantages and disadvantages are distributed between information organisation and retrieval.

Advantages include the following:

- One of the primary objectives of vocabulary control is to ensure that subjects are described consistently by both the indexer and the information seeker. If controlled vocabulary is used, subject description tends to be more consistent since the indexer is guided to acceptable terms. If, for example, a document deals with the rise in food prices, the thesaurus (or other form of vocabulary control) may indicate that the term to be used for subject description should be 'inflation' rather than 'price increases'.
- Vocabulary control also plays an important role in the control of synonyms and homographs, and is thus able to limit the scattering effect of related information.
- Controlled vocabulary deals with the problems associated with spelling variations, word forms, and so on that are experienced with natural language (refer to the previous section for more information). The user in particular needs to consult additional reference works to identify relevant terms that can save a lot of time.
- Controlled vocabulary can also indicate relationships between terms, especially generic relationships, such as 'Burmese cats' that fall under a sub-category of 'cats'. In this way an exhaustive search (i.e. a search where as many relevant and sometimes partially relevant entities are retrieved) is made possible, because the user is led to the terms in a logical manner.
- Since controlled vocabulary can reduce the number of terms which can be used for indexing, it is easier to choose an appropriate term, especially when retrieving information.

The following are disadvantages of using controlled vocabulary:

- Controlled vocabulary can never be as up to date as natural language. It is far easier to include new terminology in titles and abstracts than in lists of controlled vocabulary.
- It takes a great deal of time and money to maintain controlled vocabulary.
- Although a lot of time is saved in the retrieval process, it requires more time to prepare the inputs for the ISR, in other words to translate the concepts.

- Controlled vocabulary often does not keep up with changes in terminology and therefore with reality – that is the terms used in real life.
- It requires special training to use controlled vocabulary. The relationship between broader and narrower terms, for example, is not always apparent to novices.

Quality of Verbal Subject Description

The quality of verbal subject descriptions depends *inter alia* on

- accuracy of the intellectual content's description
- consistency of the subject description
- exhaustivity of the subject description
- specificity of the subject description

It is therefore advisable to work according to a policy for verbal subject description, such as an indexing or abstracting policy (see chapter 5).

Conclusion

Different types of verbal subject description or a combination of these can be used to describe the intellectual content of an entity. Verbal subject description stands in contrast to notational descriptions. When deciding on a type of verbal subject description (e.g. indexing terms, titles or abstracts) one should consider strengths and weaknesses, and their potential impact on information retrieval, as well as the arguments about natural language and controlled vocabulary.

BIBLIOGRAPHY

Aitchison, J, Gilchrist, A & Bawden, D. 2000. *Thesaurus construction and use: a practical manual.* 4th ed. London: Aslib.

Blair, D. 1990. *Language and representation in information retrieval.* Amsterdam: Elsevier Science Publishers.

Chu, H. 2003. *Information representation and retrieval in the digital age.* Medford, NJ: Information Today.

Cleveland, DB & Cleveland, AD. 2001. *Introduction to indexing and abstracting.* 3rd ed. Englewood, Colo: Libraries Unlimited.

Harter, SP. 1986. *Online information retrieval: concepts, principles, and techniques.* San Diego: Academic Press.

Hjørland, B & Nielsen, LK. 2001. Subject access points in electronic retrieval. *Annual Review of Information Science and Technology*, 35:249-298.

Ingwersen, P. 1992. *Information retrieval interaction.* London: Taylor Graham.

Lancaster, FW. 2003. *Indexing and abstracting in theory and practice.* 3rd ed. London: Facet Publishing.

Soergel, D. 1985. *Organizing information: principles of data base and information retrieval systems.* San Diego: Academic Press.

Taylor, AG. 2004. *The organization of information.* 2nd ed. Westport, Conn: Libraries Unlimited.

7 BOOK INDEXING

Ina Fourie and Marlene Burger

Abstract

The intricacies of book indexing are discussed. These include the clarification of concepts, steps in indexing, the choice of index entries, cross-references, locators, lead-in entries and layout. Indexable matter and what to exclude, working according to an indexing policy and the editing of the index are also considered. Please note that the technical presentation of a book index is discussed in chapter 15.

Introduction

Book indexes are essential to providing micro-level access to the content of an entity such as a book, report, encyclopaedia, yearbook or a hypertext document (i.e. access to the finer detail and pieces of information covered in a document). The importance of book indexing should be obvious from the following incident on which Cleveland and Cleveland (1990:142) report: 'Several years ago a British politician introduced a bill that would have denied copyright and imposed a fine on a publisher who issued a book without an index, but the bill failed.'

According to Lipetz (1989)

> An index can be made very short or very long, very simple in its design or very complex; but one thing an index cannot be is totally exhaustive. As most of us realize and accept – sometimes only intuitively – any text or collection can be indexed to any degree or length one might imagine, and one would still, always, be able to find more entries or aspects that could logically be added to the index. Of course, one tends not to incorporate entries in an index unless they seem "useful". But deciding what to include, how much to include, how to include it, are problems in indexing that always cry out for rational, exact answers that generally elude us. Still, we know that if there are any answers to be found, they are somehow connected to the question of usefulness, or probable usefulness, of the index that we might produce.

Definition of Concepts

Book indexing is the process whereby an index to the content of an entity is created. It is often seen as an art form because it requires creative as well as writing skills. The book index that is created in the process is a systematic guide to the content of an entity. It can

even be created for fiction. It serves the same function as an IR system: it organises the content of a specific entity in such a way that users can retrieve information from the entity when needed.

Against this background of the *function of an index* (to provide an efficient means of tracing information), the indexer should

- identify and locate relevant information
- discriminate between information on a subject and passing mention of a subject
- exclude passing mention of subjects that offers nothing significant to the potential user
- analyse concepts treated in the text so as to produce a series of headings
- ensure that terms used in the index are appropriate to users of the index, so that they will
  - quickly establish presence/absence of information on specific subjects in an unfamiliar work
  - quickly retrieve information on a remembered item in a known/partially known work
  - quickly identify appropriate entities in a collection
- indicate relationships between concepts
- group together information scattered by the arrangement of the entity or collection
- synthesise headings and subheadings into entries
- direct users seeking information under terms not chosen for index headings to headings used by making cross-references
- arrange entries into a systematic and helpful order

The indexer provides access to the content of the entity by

- conceptualisation (finding out what the document is about, and identifying important points)
- selecting concepts that are worth listing in the index (these are the entries in the index)
- translating the index entries into vocabulary (words) that will probably be used by the readers of the entity, and that will satisfy their need to find specific items of information

When dealing with book indexes, one should also understand a number of other key concepts. An entity (or document), for example, is the medium where a message is recorded. It is applicable to print as well as non-print media such as microfiche, pictures and audio, as well as other human-made objects. The purpose of an entity is to convey information.

An entry (or index entry) in a book index consists of a main heading and all other information. An entry can comprise a combination of single characters (e.g. B12 in a book about vitamins), a word (e.g. oranges), a compound word (e.g. student academic performance) or a phrase (e.g. for legal advisors). It can also include the names of organisations and institutions (e.g. University of South Africa), publications (e.g. *South African Journal for Higher Education*) or people (e.g. Nelson Mandela).

An entry can also include one or more subheadings, cross-references and locators. The layout and use of capital letters or lower case vary considerably. One has to decide on a format and standardise accordingly. The following is an example of an entry:

Book indexing
  policy 40
  purpose 39-40
  steps 40-43 *see also* Web indexing

Main heading: book indexing
Subheadings: policy, purpose, steps
Cross-reference: *see also* Web indexing
Locators: 40, 39-40, etc.

Sometimes the main heading is also referred to as an access point, entry, indexing entry, heading or subject heading. The subheading is sometimes also referred to as the sub-entry or modification. The subheading is subordinate to the main heading and it includes a specific aspect of the main heading. A subheading may comprise a combination of single characters, words, compound terms, phrases, et cetera. A locator may also be called a reference locator. A locator may consist of page numbers or paragraph or line numbers. It indicates where information about a main heading or subheading may be found.

A book index also consists of lead-in entries. They reflect the concepts that readers could use when searching for information. Lead-in entries are also called non-preferred entries. *See* references are used to guide the user from the lead-in entries to the entries preferred by the index. For example:

Back-of-book indexes *see* Book indexes
ANC *see* African National Congress

See also references are used to guide the readers to related terms which may also contain useful information. For example:

Indexers *see also* beginners; contracts; freelance work; in-house indexes

Reasons for Book Indexes

Book indexes are created for the following reasons. These also show the importance of spending sufficient time and effort when creating indexes:

- To help readers find specific facets of information about a subject as quickly as possible.
- To give access to the content of an entity. For example a textbook is of little use if you have to read through the entire book to find a specific item of information.

- To indicate subjects or items that might be important to the reader; although there might be exceptions, information mentioned in passing is normally not included in the book index.
- To give access to the terminology of the entity as well as the terminology that the user might prefer. Cross-references are normally used to guide readers from the terms or concepts that they have in mind to the terms and concepts that are used as indexing terms.
- To group terms for related matters together; this is done by using subheadings.
- To create entries for concepts covered in the entity and to combine them in such a way that users find it easy to access the information.
- To give an accurate reflection of where the information can be found (e.g. by providing the correct page numbers).
- To present a systematic series of entries so that it is easy to find specific information.

Components of a Book Index

Apart from a list of alphabetically arranged, systematic entries, a good book index should also have an introductory note, which explains the components of the index. Users must know what is covered in the index and what is excluded. Any special characteristics of the index that may affect its use should be explained. Generally only one or two features are explained, taking up only a few lines. For large multi-sequence indexes with complicated locators, abbreviations and typographic indicators, several paragraphs of explanation and guidance may be necessary. One or more of the following features may be explained:

- coverage (parts of the text covered by the index and which not)
- order of entries, if not obvious (e.g. 'broken order', 'institute' for 'institution', 'Mac' and 'Saint' forms)
- make-up of complex locators (e.g. date, volume, part, page)
- use of typestyles such as italic and bold to indicate illustrations, or locators for major references
- special use of capitals
- use of abbreviations in headings (e.g. ASAIB, LIASA) or for locators (e.g. fig for figure, n for notes)
- treatment of multi-element names (e.g. names with prefixes)
- existence of variant forms and spellings of names (e.g. Phillips and Philips; De La Mare and Delamare)
- omission or transposition of definite and indefinite articles (e.g. the, a, an)
- cross-referencing (its purpose)

The following are examples of introductory notes:

- Entries are in word-by-word order, which takes account of spaces and hyphens between words, therefore 'fishing-rods' will precede 'fishmongers'. Locators followed by *n* indicate note numbers, for example 65*n*3.
- The index covers the Introduction, Chapters 1 to 19, and the Chronology. Titles of publications are italicised, for example *Discovering South Africa.* Titles beginning with 'A', 'An' and 'The' are entered under the next word, for example *Sweet Essence of Hom*e, The. Personal names are entered according to their national or cultural customs; most Western names are placed under surname. Customs vary for multiple-element names, so if one of these is not where you expect to find it, look for it under another element.
- NOTE: The index is alphabetized in word-by-word order, for example data protection is filed before databases. Numbers in headings are filed as though spelled out, for example *W3*C is filed as *WthreeC*. Prepositions in subheadings are ignored in filing. Page numbers are given in full form, e.g. 124-125. Page numbers followed by *bib* refer to bibliographical references. Where more than one page number is listed against a heading, page numbers in bold indicate major treatment of a subject. The following abbreviations are used in headings in the index: ASI for American Society of Indexers; AusSI for Australian Society of Indexers; SI for Society of Indexers. [This introductory note appears at the beginning of the index to Booth, PF. 2001. *Indexing: the manual of good practice.* München: Saur.]

The other components have been already been introduced and will be mentioned throughout the remainder of this chapter. They are: entries (main entries, lead-in entries, preferred terms), sub-entries, cross-references and locators.

How to Index a Book

Indexing a book is a difficult process, which can, however, be mastered with practice and a critical outlook. Booth and Piggot (1988:7) explain the problems encountered in book indexing as follows: 'Secondly, and entailing much more difficult, but vital decisions, it is the indexer's duty to summarize in a succinct entry, the gist of a page, paragraph, or other section devoted to a single idea, and also to make entries for all significant topics mentioned in that section.'

The indexer should always think of the readers – the people who will use the index. Indexers usually do not know the readers and thus get no feedback about the usability of their indexes. The indexer often has to provide for the needs of a variety of readers, which makes it even more difficult. Encyclopaedias, for example, are often used by laypeople, children and experts; they must therefore contain scientific terms as well as terms for the layperson.

Information One Needs before Starting on the Index

When preparing a book index one needs the following information from the publisher or editor:

- length of the index (i.e. how many pages)
- which house style is preferred (if any)
- whether uniformity with other publications in the same series is necessary

- what supporting material is available to the indexer (e.g. the table of contents, appendices, illustrations and changes to the page drafts of the entity, all or any of which may influence the index)

Apart from the above, which refers mainly to the physical presentation of an index, one must decide on the following matters before starting the indexing process (some made in collaboration with the publisher, others made by the indexer):

- Choice of headings (i.e. which concepts to use); terminology (synonyms, offering a choice of one; linked terms like antonyms (e.g. 'good' and 'evil'); homographs with qualifier added (e.g. 'races (ethnology)' and 'races (sport)').
- Form of headings/subheadings, for example singular and plural, orthography, phrases, prepositions.
- Scope and use of headings/subheadings: sometimes scope notes may be necessary, however rare in book indexes; indentation (one or more indents).
- Proper names, which include personal names, corporate bodies and geographical names.
- Titles of entities, for example books, periodicals, music, art, films, literature.
- Locators: which to indicate and how (e.g. page numbers, volumes, columns, files, footnotes).
- Cross-references, that is *see* and *see also* references.
- Arrangement: word-by-word or letter-by-letter; headings beginning with the same term, for example

  milk
  cows'
  goats'
  milk allergies
  Milk (report)
  Milk Research Board

(The above example includes a term with subheadings, a term with a qualifier, a compound term, and the name of a corporate body. The inclusion and positioning of cross-references should also be considered.)

- Presentation of the index: printed or electronic; layout (e.g. spacing, indentation, run-on text, columns, typography).

Try to get hold of as much supporting material as possible.

Using Standards and Aids

When compiling a book index, it is important to take note of the guidelines offered in standards on indexing. They include

- British Standards Institution. 1996. *Information and documentation – guidelines for the content, organization and presentation of indexes*. BS ISO 999:1996. London: British Standards Institution. (This standard offers guidelines – applicable to all kinds of print,

non-print and electronic materials – for the content, organisation and presentation of indexes, with illustrative examples. Basic principles and practice, on which indexes can be built, are its main concern. It does not deal with detailed indexing procedures or with the automated generation of indexes.)

- *Recommendations for the preparation of indexes to books, periodicals and other documents*. 1988. London: British Standards Institution. (BS 3700:1988).
- *Recommendations for examining documents, determining their subjects and selecting indexing terms*. 1984. London: British Standards Institution. (BS 6529:1964).

More information on standards can be obtained from the website of the International Organization for Standardization at http://www.iso.ch/iso/en/ISOOnline.frontpage/

Steps in Book Indexing

Book indexing consists of the following steps:

- conceptualisation, which means analysing an entity to determine what it is about
- selection of concepts or items worth an index entry
- translation of concepts into index entries (e.g. main headings, subheadings)
- indication of the location (e.g. page numbers) where the information appears
- creation of lead-in entries (*see* references) that guide readers from familiar terms to those used in the index
- creation of cross-references (*see also* references)

What to Index

Certain parts of an entity contain 'indexable matter' – information that might be worth indexing. The author, book editor, publisher or indexer must decide on the sections of an entity that will be indexed. Must the entire entity be indexed, or only the major sections, or only certain sections? When making such decisions, the potential readers should also be considered.

Sections of an entity which are usually indexed include

- the text (this can include anything from the main sections to only selected sections)
- Introduction
- glossaries
- notes (footnotes, endnotes)
- addenda
- illustrations
- appendices

Fetters (1994:14) is of the opinion that word lists do not need to be indexed. References (such as bibliographies, reading lists, further reading and lists of works consulted) are only

sometimes indexed. The following are normally excluded from the indexing process (but not always):

- title page
- name of the person to whom the entity is dedicated
- table of contents
- synopsis of chapters

The type of entity will also determine what is considered to be 'indexable matter'. For example an academic text will be indexed in detail, whereas an index for an annual publication or a journal might only contain the main articles instead of the news items and items of passing interest.

Studying the Physical Entity

One needs to study the physical entity to determine its contents. The subject content must thus be determined. This will be determined by the decision on what is to be considered 'indexable matter'.

Initially, one should make a preliminary study of the entity and skim through the introduction to get a general feel of the text, the potential readers and their information needs. Cleveland and Cleveland (1990:129) advise indexers to write a summary of 500 words of the entity; if an indexer cannot write such a summary, he or she will probably have problems indexing the entity.

The most important matters or components that one has to cover in the initial examination are

- title
- contents page
- introduction
- concluding paragraph and summary
- list of illustrations
- word list
- headings
- words and phrases highlighted in the text

After the provisional examination of an entity, read it a second time in more depth. One can begin to make notes of the main headings and note the major and the less important terms and concepts. It is very important to note new and unique information. (This is difficult if one is not really familiar with the topic.) Each page and each sentence must be considered during the indexing process. Terms which might be particularly important should be underlined. Not every word in a sentence is used for indexing, however. In the sentence 'Children are like rays of sunshine', only the term 'children' is important as an index term. Terms in bold and italics must always be considered. Otherwise one need only take note of particular words if mentioned in a fair amount of detail. Hence the key question: What is a fair amount of detail?

When deciding whether it is worth making an index entry or not, keep the readers' needs in mind and the amount of information in relation to the subject as a whole.

There are no specific guidelines for the number of index entries which may be selected for an index. Sometimes the length of an index is determined by the publisher – this might mean that the indexer can only compile a brief index.

Keep an open mind when studying an entity to determine the concepts for indexing. Indexers must index the content of an entity and not their own ideas about what the entity should cover.

Translating the Content and Creating Entries

Decide on main headings, subheadings and cross-references for all entries, always keeping the readers' needs in mind (e.g. the questions users might ask, the subject they want information about and the words they might use).

Many sources can assist the indexer in the choice and formulation of main headings and subheadings:

- standards
- dictionaries
- word lists
- encyclopaedias (general and subject-oriented)
- guides
- classification schedules
- thesauri in the subject discipline and related disciplines
- house style manuals
- guides to the use of English
- other experienced indexers

The website of the ASI offers an extensive list of links to such sources: http://www.asindexing.org

Style for Interpreting Entries

Some publishers have a house style. If not, the indexer can decide on his or her own style. Standards such as the following can give valuable guidelines:

- *Recommendations for the preparation of indexes to books, periodicals and other documents*. 1988. London: British Standards Institution (BS 3700:1988).
- British Standards Institution. 1996. *Information and documentation – guidelines for the content, organization and presentation of indexes*. BS ISO 999:1996. London: British Standards Institution.
- *Chicago manual of style*. 1993. 14th ed. Chicago: University of Chicago Press.

The goal is to formulate the entries as clearly and understandably as possible. The direct format is therefore generally chosen when more than one word is used, for example 'academic libraries' rather than 'libraries, academic'. Exceptions are made for individual

names (e.g. Mandela, Nelson, rather than Nelson Mandela). The opposite is recommended if it helps to group related information together. For example:

- alphabetising, letter-for-letter (instead of letter-for-letter alphabetising)
- alphabetising, word-for-word (instead of word-for-word alphabetising)

Indexes usually consist of two or more columns per page and are printed in a font smaller than the rest of the entity. One column per page takes up too much space.

Generally two basic styles are used to construct a book index: the indent style and the run-in style (or run-on style). The indent style gives a clearer construction, while the run-in style is used primarily to save space. The following is an example:

Indent style	**Run-in style**
dogs breed from 180-185 breeding clubs 242-248 obedience training 275-281	dogs: breed from 180-185; breeding clubs 242-48; obedience training 275-281

Special typography can also be used, such as italics to indicate the titles of publications or entries for illustrations. Bold can be used to indicate page numbers with key information about a subject. For example:

ASAIB newsletter 225
children's books 8, 14, **118-125**, 138, 177
papyrus 14, *82*, 91, **95-99**, *107*

Formulating Main Headings and Subheadings

There are no precise guidelines for formulating headings or subheadings. Individual indexers will probably have different entries for the same entity. However, there are guidelines that will help with the indexing task when it comes to the following points for example:

- personal names
- geographical names
- names of organisations
- abbreviations and acronyms
- international characters
- capital and small letters

The following tips might also help with the formulation of main headings and subheadings:

- Verify unknown terms.
- Verify inconsistent use of terms by the author.

- Use subheadings if there are more than five pages of references that must be included with the main heading. For example:

```
Kenya 34-38, 48-52, 60, 66, 70-71, 77-85
Kenya
 British period 34-38, 60
 Early history 66
 economy 48-52, 60, 70-71
 Kenyatta administration 77-85
```

- An incomplete entry, such as 'Disraeli', must be verified and reformulated so that it reads 'Disraeli, Benjamin, 1st Earl of Beaconsfield'.
- Keep a balance between key information and less important information.
- Do not index negative information (e.g. do not make an entry for Nelson Mandela if the text indicates that Mandela's presidency is not covered in the study).

Not all entries can (or should) be exactly as they appear in the entity. Sometimes indexers use reference works to adapt index entries (e.g. 'Disraeli' can become 'Disraeli, Benjamin, 1st Earl of Beaconsfield').

Indicating Relationships between Entries

Entries in a book index are arranged in such a way that the relationship between main headings and subheadings is very clear:

```
learn to index
  bibliographies, 24-26
  correspondence courses, 45
  self-training, 45-47
```

Recording Entries

Write the entries on a card or use a computer program such as Macrex or CINDEX to generate the entries. Make separate entries, irrespective of the method used. This simplifies the alphabetical sorting of the main headings and subheadings.

Locators

The position where information about a particular heading or subheading can be found, usually given as page numbers, must be correctly indicated. The page numbers are sometimes separated from the heading or subheading by a comma. When a heading or subheading has more than one page number, the page numbers must be separated by a comma. A hyphen can be used to indicate a range of page numbers (e.g. 50-58, 102-110). For example:

```
indexing
  exhaustivity 43-45
  specificity 24, 34
```

Making Cross-references

There are two different types of cross-reference:

- *see* references
- *see also* references

See references guide the user from the term not used to the term preferred by the book index. For example:

```
American National Standards Institute see ANSI
distance teaching see distance education
```

See also references guide users from an index entry to another related entry. There may be more than one *see also* reference for an entry; these are then arranged alphabetically and separated by commas. *See also* references can follow a main heading (i.e. after the page numbers if there are any), or they can be listed as a group of subheadings. For example:

```
libraries see also academic libraries, public libraries
indexing
  as art 56
  history of 56-67
  see also methods of indexing
```

General *see also* references are used to ensure that users consult the general and specific entries. For example:

```
churches see also names of specific churches
universities see also names of specific universities
```

When making a cross-reference ensure that the entries referred to really exist. Avoid entries that run in circles. For example:

```
libraries see academic libraries
academic libraries see libraries
```

Arranging Entries

The entries can be arranged alphabetically and there are two options: word-by-word or letter-by-letter. For example:

Letter-by-letter	Word-by-word
New Amsterdam	New Amsterdam
Newark	New Brunswick
Newberry	New England
New Brunswick	Newark
New England	Newberry
Newfoundland	Newfoundland

How to Work according to a Policy

Although there are no fixed rules for book indexing, one needs to clarify a number of issues before starting on the index. Consistency is essential. The indexer can work according to his or her own policy or according to the policy of the publisher, if specified. A policy lays down the rules and guidelines, which must be followed.

It is particularly important to be consistent about the following:

- level of detail provided
- level of vocabulary
- form of terms
- style of entries
- layout
- presentation
- spelling
- punctuation and spacing
- use of capital letters
- word order if there is more than one word
- order of headings and subheadings
- use and form of cross-references
- word forms for preferred terms
- forms of locators (e.g. references to page numbers)

There are many standards, textbooks and other sources that one can consult for information on these matters. See also chapter 15.

Specificity and Exhaustivity of Entries

Specificity of indexing refers to the degree to which an index entry matches the exact meaning of the subject concept (Cleveland & Cleveland 2001:259) or the vocabulary used in the entity. Exhaustivity refers to the degree to which the subject content of an entity is reflected in the entries. It also indicates how much information about a subject can be retrieved and how easily.

The more specific indexing terms are usually assigned (e.g. 'distance teaching' rather than 'teaching'). If a book deals with the purchasing policy of public libraries and academic libraries, the heading will be 'public libraries' and 'academic libraries'; these are preferred to the more general term 'libraries'. A cross-reference is made between 'libraries' and the

more specific terms, namely 'public libraries' and 'academic libraries'. Sometimes the more general term 'libraries' is also assigned if it will benefit the user. If a book deals briefly with a series of related terms, the more general term will be suggested (e.g. 'fruit' to replace 'oranges' and 'peaches').

See also chapter 4 where these two concepts are discussed in more detail.

Synonyms, Antonyms and Homographs

The policy must include guidelines to the handling of synonyms, antonyms and homographs.

Synonyms are terms with more or less the same meaning, such as trespasser/criminal or inflation/price increases. Antonyms are terms that have the opposite meaning, such as pass/fail. Such terms must be approached in a special way regarding headings and subheadings:

- One term is chosen as the preferred term for synonyms. The locator is indicated here. References are made from the non-preferred term. For example:

```
inflation 75-80, 86
price increases see inflation
```

Synonyms and near-synonym words can also be linked (e.g. libraries and librarians):

- For antonyms one of the preferred terms is chosen. A reference is made from the non-preferred term to the preferred term. Alternatively, an entry can be made for both terms. For example:

```
employed and unemployed 56, 60-65
unemployed and employed 56, 60-65
```

- For homographs the meaning of a term is indicated in brackets. For example:

```
Mercury (planet)
mercury (metal)
```

Form of Headings and Subheadings

The form of headings and subheadings must be standardised; usually nouns are chosen. Pronouns start with capital letters, but other headings use small letters. Descriptive phases can use subheadings. A heading usually starts with a noun, but sometimes an adjectival noun phrase is more specific and gives a more realistic reflection of the term preferred by the user.

Two related terms can also be combined. For example:

```
colleges and universities
libraries and librarians
```

The terms will not be linked if the author differentiates between the terms, or if he or she provides a large amount of information about one of the terms, but not the other. Here a main heading must be made for each concept. The two entries must be linked with a *see also* reference.

Singular or Plural

The plural form is preferred for countable nouns (i.e. 'how many?'), such as libraries, dogs or books. Nouns for uncountable nouns (involving mass; i.e. 'how much?') use the singular form, such as gold and water. The singular form is also used for abstract concepts, processes, attributes, quantities and characteristics.

Cross-references

There are no definitive guidelines for making cross-references apart from the following:

- differences in spelling (e.g. alphabetizing/alphabetising)
- words and names in foreign languages
- archaic words
- synonyms and antonyms
- genus to subclass (e.g. fruit/citrus fruit/oranges)
- from subclass to genus (e.g. oranges/citrus fruit/fruit)
- name variations
- abbreviations and acronyms (e.g. ANC/African National Congress)

Editing the Text

The index must be thoroughly edited after all the entries have been made. Check for the following:

- consistency of entries
- duplication of entries
- proper and correct cross-referencing
- spelling (this includes consistency of spelling)
- correct page numbering (or other forms of locators)
- correct alphabetising
- adherence to the standard procedure for filing according to the ALA filing rules

It is very useful to have a second person help check the technical points. See chapter 15 for more detail.

Preparing the Final Index

The final index can be prepared after the draft copy has been proofread and checked. Do not start this process before all the errors have been corrected. The index should be checked

and rechecked thoroughly. Unfortunately, publishers do not always give indexers enough time to check their work.

And finally: always keep a backup copy of the index!

Indexing Software

Traditionally indexers indexed books manually on cards. Although cards may still be used, it is much easier to use an indexing program. Nowadays various single-purpose programs are available for indexing, such as CINDEX, Macrex and SKY Index. These programs carry out certain mechanical tasks, leaving the indexer to provide the intellectual input. Embedded indexing and automatic indexing are discussed in chapters 9 and 10 in detail. The following functions are supported by indexing programs:

- automatically displaying a group of entries that contain a common term or phrase
- checking cross-references
- sorting and arranging terms (including variations in the sorting sequence and arrangement styles)
- generating the printed layout chosen by the indexer
- checking indented and run-on entries
- allowing the indexer to determine spaces for subheadings and sub-subheadings

Word or text processing programs can also be used. Even if these have an indexing module, they are not as effective as single-purpose programs. Other possibilities include embedded indexing (Microsoft Word, Adobe PageMaker and markup languages such as SGML, HTML and XML) and tagging. Input is eased by the provision of a range of sophisticated facilities for

- repeating, rotating, merging, sorting, and styling, of the individual entries
- reviewing and amending the index during compilation
- checking the effectiveness of cross-references
- displaying and producing the final index in the required format (e.g. hard copy, disc, camera-ready copy, marked-up files for electronic transfer)

For more information about these programs, you may like to visit the following websites:

Idxtools http://www.sid.com.ac.uk/ndv207/idxtools.htm

CINDEX http://www.indexres.com

Macrex http:///www.macrex.com

SKY Index http://www.sky-software.com

Master Indexing http://www.mindexer@interconnect.com.au

Although computer programs can help simplify the book indexing process and can improve productivity, human intellectual effort is still needed (Wellisch 1995:247). This statement is

still valid today. Not even the most sophisticated automatic indexing system has been able to produce an acceptable book index.

Conclusion

Book indexing is an important source of freelance work and is gaining in popularity. Good indexers with knowledge of book indexing are in demand with publishers. If one is planning to make a career of book indexing, various courses offered by indexing societies should be considered (for more information, see chapter 31 on training opportunities). The best way to keep up to date with developments in the field of book indexing is to regularly visit the various websites of book indexing societies. A number of associations focus on book indexing, for example

- The American Society of Indexers (ASI): www.asindexing.org
- The Indexing and Abstracting Society of Canada (IASC): www.indexingsociety.ca
- The Australian and New Zealand Society of Indexers (ANZSI):www.aussi.org
- Society of Indexers (SI) (UK): www.indexers.org.uk
- Association for Southern African Indexers and Bibliographers (ASAIB):www.asaib.org.za

BIBLIOGRAPHY

Bell, HK. 1992. Indexing biographies and other stories for human lives. Society of Indexers Occasional Papers on Indexing, no. 1). London: Society of Indexers.

Blakewell, KGB. 1987. Reference books for indexers. *The Indexer*, 15(3):131-140.

Booth, PF & Piggot, M. 1988. *Choice and form of entries*. (Society of Indexers: Training in indexing – Unit 2). London: Society of Indexers.

Borko, H & Bernier, CL. 1978. *Indexing concepts and methods*. New York: Academic Press.

Chicago manual of style. 1993. 14th ed. Chicago: University of Chicago Press.

Cleveland, DB & Cleveland, AD. 1990. *Introduction to indexing and abstracting*. 2nd ed. Englewood, Colo: Libraries Unlimited.

Cleveland, DB & Cleveland, AD. 2001. *Introduction to indexing and abstracting*. 3rd ed. Englewood, Colo: Libraries Unlimited.

Fetters, K. 1994. *Handbook of indexing techniques: a guide to beginning indexers*. Rochester, NY: Fetters InfoManagement Co.

International Organisation for Standardisation (ISO). 1996. *Information and documentation – guidelines for the content, organization and presentation of indexes.*

Knight, GN. 1979. *Indexing, the art of: a guide to the indexing of books and periodicals*. London: Allen & Unwin.

Lancaster, FW. 2003. *Indexing and abstracting in theory and practice*. 3rd ed. London: Facet Publishing.

Lipetz, B-A. 1989. The usefulness of indexes, in *Indexing: the state of our knowledge and the state of our ignorance: proceedings of the 20th Annual Meeting of the American Society of Indexers*, edited by BH Weinberg. Medford, NJ: Learned Information:111-120.

Mulvany, NC. 1994. *Indexing books*. Chicago: University of Chicago Press.

Recommendations for the preparation of indexes to books, periodicals and other documents. 1988. London: British Standards Institution. (BS 3700:1988).

Thatcher, VS. 1995. *Indexes writing, editing, production.* Lanham, Md: Scarecrow Press.

Wellisch, HH. 1995. *Indexing from A to Z.* 2nd ed. New York: HW Wilson.

Wellisch, HH. 2000. *Glossary of terminology in abstracting, classification, indexing and thesaurus construction.* Medford, NJ: Information Today.

8 EMBEDDED INDEXING FOR CREATING BOOK INDEXES

Ethleen Lastovica

Abstract

The indexing of electronic text is discussed. Indexers use a variety of software that has embedded indexing modules. Specific reference is made to the indexing of the B*usiness blue book of South Africa*.

Introduction

Embedded indexing is a paperless process that has been defined as the inserting, or embedding, of index entries into an electronic document (a computer file) with the software that was used to create that document. The process as described in this chapter is done manually by an indexer, unlike automatic indexing where keywords are automatically assigned to the text by software according to predetermined guidelines. Once subject entries have been embedded in the text, an index of these terms can be compiled to indicate on what page the terms occur (Mulvany 1999).

By considering the indexing of the *Business blue book of South Africa*, the preparation, processes and publication of an index for a conventional book using the techniques of embedded indexing are discussed.

Hiking on the Otter Trail

Embedded indexing can be compared to a hike on the Otter Trail. The guidebook for the five-day walking trail describes the terrain as one of undulating hills with steep cliffs and a magnificent coastline intersected by rivers entering the sea. This creates an image of happy hikers ambling through a pastoral scene with spectacular views. One needs to consider, though, that the cliffs need to be scaled in order to reach the best vantage point for the magnificent views, and that one might have to swim flooded rivers to reach the other bank.

There is an increasing amount being written about embedded indexing and its pros and cons, but until one is confronted with the prospect of undertaking this form of paperless indexing, the words of advice, and warnings about its pitfalls, are not fully appreciated.

Definition / Software

Embedded indexing is done by indexers who select index terms for inserting, that is embedding, into the text of a publication when it is still in electronic form. The software that was used to prepare the electronic document is used. The index can be compiled

once the whole document or publication has the embedded terms in place. Various word-processing and desktop publishing programs, such as Microsoft Word, WordPerfect, Ventura, FrameMaker, PageMaker and QuarkXPress have embedded indexing modules. Conventional book indexes are often created with dedicated indexing software – for example Macrex, CINDEX or SKY Index – but all three of these programs now include an embedded indexing facility.

Embedded indexing software is improving, but is far from perfect. It is important to be aware of the latest developments in the field by not only consulting indexing journals such as *The Indexer*, but also checking for new resources on the Web. Nancy Mulvany (1994, 1999), Jan Wright (1999, 2000), Peg Mauer (2000) and Seth Maislin (1999) are indexer/authors professionally involved in embedded indexing. While Mulvany covers embedded indexing in her 1994 publication, *Indexing books*, the reader must note that the book is 10 years old. More recently she published an article in *The Indexer* (Mulvany 1999) in which she compared the capabilities of four embedded indexing programs with Macrex, which is considered by many as the leading dedicated indexing program. The report duplicated tests done ten years earlier. It is startling to see how differently the five programs alphabetise a list of sample entries. Her conclusion was that, despite improvements, embedded indexing programs are still crude tools.

The type of person who should become involved in embedded indexing is someone who is not intimidated by computers and software, is methodical, and an expert file manager. The indexer undertaking the job should be able to stay calm as the complexity level of the task rises.

Business Blue Book of South Africa

The reality of embedded indexing will be put into perspective by describing the indexing of the annual *Business blue book of South Africa* that is published by National Publishing in Cape Town (2004). It is available as a book or on CD-ROM. The 2004 edition has 916 pages and a 151-page index consisting of approximately 29 000 entries. In the *Business blue book*, relevant facts and legislation needed to conduct business in South Africa are brought together in one volume. New legislation may be incorporated only days before going to press.

Sarah Maddox, a technical writer and indexer, and the former ASAIB representative in the Western Cape, started the embedded indexing project for the *Business blue book* some years ago. She realised the importance of sound documentation as a means of establishing standards and procedural guidelines for the team who work on the book, particularly the indexer. As successive indexers have undertaken the challenge of indexing the *Business blue book*, the procedural guide has proved invaluable, particularly as the editorial team is generally unable to provide help regarding the technicalities of indexing.

The publication is set up using the desktop publishing package QuarkXPress. The indexer uses an Apple Macintosh loaded with this software and Quark Index XTension to create the embedded index. The indexing takes about two months. It is essential that anyone undertaking an indexing assignment like this is totally familiar with the hardware and software used.

Preparations for Embedded Indexing

As with any indexing contract, speed of indexing is crucial. Computers with high capacity are essential because the files for an entire book created with a desktop package are usually large. A modem providing a link to the Internet and e-mail is essential. When quoting for the job one must not forget to include costs for online time because large files take a long time to download and return.

Discussions with the editor and production director must establish a sound understanding of what is required, and the administrative procedures for ensuring the safe electronic transfer of files between editor and indexer.

Conceptual Approach

If indexing an annual, it is advisable to understand the conceptual approach that previous indexers have had, and to become familiar with the terminology they used by studying earlier indexes to the work.

Each chapter of the *Business blue book* is a separate file. It is vital to work only on the current version of a file, which means the editor and indexer cannot work simultaneously on the same file. A progress chart, or spreadsheet, must be meticulously maintained to keep track of who has the current editing rights to a file. The spreadsheet should list chapter title and filename, and include columns for date received, date completed, index tested, date returned and comments.

Procedures

When the editing of a section or chapter of the *Business blue book* is complete and the pages of that section have been typeset, the QuarkXPress file is e-mailed to the indexer for indexing. The chapters may be sent in random order. Besides saving the incoming files on the computer on which the book will be indexed, it is advisable to keep secure originals of these files apart from that computer. 3,5 inch double-sided high-density diskettes (stiffies), CD-ROMs, a zip drive or a flash drive could be used for storing backups. Files should be saved frequently while working on them, perhaps every 10 minutes. At least once a day a backup must be made on one's preferred storage medium and kept apart from the computer in case of system failure. This may seem over-cautious, but should the working file become corrupted it would be time-consuming to reconstruct the index. Embedded indexing generally takes two to three times longer to do than conventional book indexing does with specialised indexing software.

Besides receiving files electronically, page proofs of the chapters are delivered to the indexer. Initially, some might find it easier to mark up the proofs with indexing terms that one intends embedding in the electronic file, but with a little practice it saves time to work directly on the file. The electronic text from earlier editions of a publication retains the embedded indexing terms that were originally inserted into it. The difficulty with working from proofs is that the indexer cannot see what has previously been indexed because, in print, embedded indexing terms are hidden text. Only new or updated text, which the editor

indicates by underlining, needs to be indexed. This is what publishers may consider as one of the advantages of embedded indexing.

Embedded indexing is cumbersome. Until the entire volume has been completed, only separate indexes for each file or chapter can be generated – the amalgamated index for the whole work is only created as the last process before going to press. It is therefore not easy to check, or even remember, where else a topic occurred and how it was indexed. This leads to inconsistency in the index. If indexing an annual, or a new, edition of a work, it may be possible to solve this problem to some extent by consulting the index of the previous volume. If that index was set in very small font, it helps to have an enlarged photocopy of it to make the text more legible. Indexers born into the computer age who have a large computer screen might find it just as easy to have the electronic file of the earlier index open while indexing the new edition.

Indexing Decisions

Many hours can be wasted considering how to deal with previous indexing decisions that one disagrees with. It is not good indexing practice to change the existing indexing terms in the current file or chapter if all other occurrences in the publication are not changed as well. Writers on embedded indexing generally consider changes like this to be impractical because the process would be time-consuming and confusing, particularly as one might not have the entire publication on file at any one time. The best approach is to follow the indexing decisions made for earlier editions. (Embedded indexing is about compromise.) When too many inconsistencies or inaccuracies build up in an index to a long-running publication, a stage will be reached when the document must be completely re-indexed.

Procedure for Indexing an Electronic Document

- Receive file from the editor.
- Update the index by embedding additional indexing terms into the text, or create a new index if the document has not been indexed previously.
- Create a test index, and proofread it.
- Go back to make changes to the embedded indexing terms if necessary.
- Return file with embedded index to the editor.

The embedded indexing terms may be invisible when the computer file is opened, but then the option to show commands should be selected. If this is a revised version of an earlier work, the editor indicates, by using a different colour, or font, what text has been added to the document. It is these segments that must be indexed. The layout defaults in the indexing software being used must be set up, including punctuation, indents and sort options.

Quark Index XTension

Using Quark Index XTension the indexer highlights the word, phrase, paragraph, section or chapter to which an index entry must refer. A pop-up menu allows the indexer to check the

index entry and to confirm or amend it, and define the level, that is a main entry (first level) or a subentry (second level). Page numbers are not assigned at this stage, only locators – the location of the index entry on a page. The tagged entry moves with the text to which it is assigned. This allows reformatting of pages after they have been indexed. Only when all the files are combined as one document, and pagination added at the end of the editing process, is the final index, complete with page numbers, created.

It is all very well to have editorial marks indicating what is new, but a problem arises because it is impossible to show what has been deleted. For this reason the indexer needs to scrutinise the text and list of entries assigned to it. If the whole section to which an indexing term referred was removed, there is no problem because the entry disappears from the already compiled list of entries. If, however, only the beginning or end of a section has changed or been removed, the start or close tags of the entry may be missing. The tagging of the entry will then have to be corrected.

When embedded indexing of a file is complete, Quark Index XTension provides an option to build an index for that file in layout style with temporary page numbers assigned to the entries. This provides the opportunity to check for inconsistencies and spelling errors which must be corrected in the text. The indexed file is then returned to the editor who is quite likely to make further changes to content or layout of the chapter. When all files are finalised, the editor generates the index for the whole book. Locators alongside the index entry are automatically replaced by page numbers, and so the index is ready for checking. Almost without exception, writers on embedded indexing consider this stage to be the most time-consuming of the embedded indexing process. For the *Business blue book* this entails checking for inconsistencies and misspellings across the 78 chapters that were separately indexed. Any errors noted must be corrected within the chapter concerned. The complete index must be regenerated if corrections to embedded indexing terms are made. This is the final stage of the indexer's commitment to the printed version of the publication.

CD-ROM

The *Business blue book* is also published on CD-ROM, and so the indexing process starts again, but this time a microcomputer and a range of Adobe software to create hyperlinks and bookmarks are used. As with the printed version, a progress chart for recording the receipt and return of files is important and regular backups of work in progress are essential. A detailed procedural guide for indexing the CD-ROM must be written to enable someone new to the job to proceed with confidence. In addition the indexer must write technical notes to explain to the readers how to use the CD-ROM.

All files that were created for the printed book are converted to PDF and then reformatted to fit the shape of a computer screen. Only an outline of the steps taken to hyperlink and bookmark the CD-ROM version of the publication follows.

Steps to Hyperlinking an Electronic Document

- PDF files (readable with Acrobat Reader) are supplied by the publisher.
- Files are processed through a template using Acrobat Exchange.

- Text is hyperlinked and bookmarked with Adobe Exchange. This will enable a reader to find a reference in the text by clicking on an entry in the bookmark menu.
- Attention is paid to special hyperlinks for certain files (e.g. e-mail links and forms).
- A simple index to subsections of each Act or chapter is prepared.
- When complete, the simple index is sent to the publisher for conversion to house style and PDF format.
- The index in PDF format is returned to the indexer for hyperlinking and bookmarking.
- Next, Acrobat Catalog, one of the programs supplied with Acrobat Exchange, is used to create an index that enables searches for words or phrases in the *Business blue book* files on the CD-ROM.
- All the hyperlinked PDF files and those created when building the search index are zipped and returned to the publisher for writing to the CD-ROM.

Task complete!

BIBLIOGRAPHY

Business blue book of South Africa – BBB. 2004. Edited by L Mitchell. Cape Town: National Publishing. 65th ed.

Maislin, S. 1999. Ripping out the pages, in *Beyond book indexing,* edited by D Brenner and M Rowland. Medford, NJ: Information Today:43-57.

Mauer, P. 2000. Embedded indexing: pros and cons for the indexer. *The Indexer*, 22(1):27-28.

Moncrief, L. 1999. Indexing computer-related documents, in *Beyond book indexing,* edited by D Brenner and M Rowland. Medford, NJ: Information Today:13-24.

Mulvany, NC. 1994. *Indexing books.* Chicago, IL: University of Chicago Press.

Mulvany, NC. 1999. Software tools for indexing. *The Indexer*, 21(4):160-163.

Wright, JC. 1999. Indexing after the millennium 1: getting the tools right. *The Indexer*, 21(1):19.

Wright, JC. 2000. The world of embedded indexing, in *Beyond book indexing,* edited by D Brenner and M Rowland. Medford, NJ: Information Today:3-12.

9 PERIODICALS INDEXING

Antoinette Kotze

Abstract

This chapter concentrates on the Index to South African Periodicals (ISAP) as an example of periodicals subject indexing database. It covers compilation, application and use by the National Library, selection and the function of its main indexing fields and discusses ways of fostering improvement. It also includes an addendum on the principles of periodical indexing.

Introduction

The Index to South African periodicals (ISAP) is an example of a periodicals subject indexing database. This chapter gives information on the following:

- ISAP as a product of periodicals subject indexing – a national indexing database
- how ISAP is compiled
- how ISAP is applied and used by the National Library and other database providers
- selection of information in non-scholarly periodicals illustrated with examples
- function of its main indexing fields
- improvements to ISAP

What is ISAP?

ISAP is a product of periodicals subject indexing that has become a national indexing database.

The Library Association of South Africa started ISAP in 1940, and it was continued by the then Johannesburg Public Library in 1945. In 1986 the then State Library acquired authority over ISAP in terms of a ministerial decision. The State Library was merged with the South African Library, Cape Town, on 1 November 1999 to form the National Library. As a product of periodicals subject indexing, ISAP represents a main function of the national bibliographic service as determined in the National Library Act 92 of 1998 (chapter 1, item 4(1)(b)(ii)).

ISAP does not reflect all legal deposit for South African periodicals, but comprises a comprehensive selection of mainly scholarly journals and technical and subject general periodicals. A limited number of popular magazines useful in schools or adult basic education are included.

As a subject index, ISAP provides access to information through natural language keyword terms, controlled thesaurus terms and abstracts, as well as a basic bibliographic

description to an article title, author and journal reference. Research and use value in the South African context form the basis for selecting subject terms for inclusion.

As a South African indexing tool, ISAP is a unique source to national periodical subject information.

How ISAP is Compiled

Contracts

The policy of the National Library in compiling ISAP is based on contracts with specialised institutions and qualified individuals.

Currently the Council for Scientific and Industrial Research (CSIR), University of South Africa (Unisa) Law Library and the National Institute for Theology and Religion (NITR) are institutions involved in indexing for ISAP. Professional subject expertise at indexing institutions can benefit ISAP by for example proposing additional periodicals for indexing that represent reliable research or new developments in a subject area. Institutions also represent a broader spectrum of institutional input in ISAP at national level.

Required qualifications for indexers include an acceptable library (or other subject) qualification with at least four years' experience in general indexing, classification and subject cataloguing. Experience in indexing of periodicals is preferred. Basic bibliographic cataloguing knowledge is required. Expertise in the use of indexing software has to be acquired. An indexer is required to have appropriate qualifications in subject areas of journals allocated for indexing or to acquire the necessary subject knowledge.

Qualified individuals undergo training in ISAP indexing and are considered for contracts with the National Library only on constant and regular delivery of trial work according to ISAP standards.

Standards, Rules and Guidelines

The following standards, rules and guidelines apply to ISAP:

- SAMARC minimum level (programmatically converted to MARC21)
- ISBD(S): International Standard Bibliographic Description for Serials
- AACRII cataloguing rules
- ISO9115:1987(E):Documentation–Bibliographic identification(biblid)ofcontributions in serials and books
- ISAP guidelines for indexing drawn up by the National Library and indexing institutions, and the ISAP thesaurus initiated by Johan van Wyk
- Q & A or Inmagic indexing program
- ISO 5963: Documentation – Methods for examining documents, determining their subjects and selecting indexing terms
- ANZI Z39.14: American National Standard for writing abstracts
- Library of Congress list of languages and language codes
- alphabet tables for transliteration of non-Roman and African languages

Standards for indexing electronic journals are not included because these journals are not indexed on ISAP as yet. The National Library is investigating this option.

Indexing Approach

Against the background of contracts, standards, rules and guidelines, indexers aim to index substantial and useful information according to subject. With selection of information, user needs feature prominently in the focus of ISAP.

Information in scholarly journals is fully indexed. Technical and subject general periodicals (e.g. *Computing SA*, and *African Security Review*) are selectively indexed, focusing on article-type presentations and appropriate subject material. Information that is useful for learners and neo-literates is selected for indexing in popular magazines (e.g. *Drum, Huisgenoot*).

The latter two groups of periodicals and magazines can include much information presented for a passing readers' market as opposed to essential information useful for research and recorded for this purpose. Information for a readers' market falls under ephemera to be excluded from ISAP.

A high priority is placed on indexers' professional insight and integrity with regard to selecting information for indexing. Regular contact with indexers is maintained to interpret new problems and reflect resulting decisions in the guidelines.

Periodical Selection

A broad spectrum of subject fields is covered by ISAP periodicals, for example natural sciences, humanities, education, medicine, technology, agriculture, theology and religion, public administration, and business and industry.

The list of periodicals inherited from the then Johannesburg Public Library (now the Greater Johannesburg Library Services) in 1945 is still in existence. Since 1986, it has been extended on a regular basis according to the focus and procedure for ISAP, accepted and developed by the National Library.

Focus

The aim pursued by ISAP periodicals has been to represent the following:

- a selection of periodicals published in South Africa, regarded as significant and useful
- important scientific, technical and subject general periodicals especially at scholarly level, appropriate for representation on ISAP
- a limited number of popular magazines with information relevant to the needs of schools and basic adult education

Extension of Periodicals Lists

It is procedure to extend the list of periodicals throughout a financial year. This ensures a continuous flow of periodical supply to indexers.

The list of periodicals will be extended for the following reasons:

- if a title is closed down
- to keep up to date with new periodical publications
- to follow up users' requests for additional titles
- to represent new areas in subject fields appropriate to the focus of ISAP

Additional Periodicals

Before being added to ISAP, periodicals are assessed against the following criteria:

- extent of support for a specific title (especially applicable with a survey on users' proposals for additional titles)
- need for representation on ISAP, considering the extent to which an additional periodical's subject area would complement existing coverage on ISAP (with the actuality of a subject area also taken into account)
- availability of funds

Proposal Procedures for Additional Periodicals

Users are invited to propose additional periodicals for indexing on the ISAP system. The following information is required to accompany such proposals:

- full periodical title and publisher
- brief motivation for its inclusion in ISAP (e.g. that it would complement existing subject coverage or support a user need)
- the first and latest issues of the relevant periodical (if it is not available at the National Library) or photocopies of the title, content and editorial pages of the named issues to give an indication of the scope of the periodical

Process of Compilation

Supply of Periodicals for Indexing

Periodical titles as stipulated in an indexing contract are supplied for indexing usually on a weekly basis, but also over longer intervals depending on availability of periodicals. The frequency of periodical publications could influence the regularity of periodical supply for indexing. Where late publication of current periodicals hinders periodical supply, retrospective issues of selected periodicals are sent instead to help maintain a regular flow of periodicals for indexing. It is mainly scholarly journals that are indexed retrospectively in this way.

Record Delivery, Quality Control and Supply of Records for Access

Indexed records are delivered on a weekly basis or over longer intervals, depending on periodicals received for indexing.

Printouts of newly processed records are checked weekly for quality control. Errors and corrections are discussed with indexers if necessary. Alternatively they receive copies of printouts to inform them of corrected minor errors – spelling mistakes and the like. It may

be necessary to resend records prior to acceptance. Corrections, other than the weekly ones, are dealt with on a continuous basis.

Following the weekly record processing and quality control, newly accepted and corrected records are loaded on the National Library's in-house ISAP database and on its Internet version, ISAPOnline, available at www.nlsa.ac.za/service_isaponline_about.html

At the same time, these records are supplied for licensed access to outside database providers such as the South African Bibliographic and Information Network (SABINET), National Inquiry Services Centre (NISC) and Potchefstroom University.

Claims and invoices for newly accepted records are paid on a monthly basis or as invoices arrive.

Application and Use of ISAP by the National Library and Other Database Providers

Access to ISAPOnline

The National Library allows non-subscribers gratis searching on ISAPOnline, where limited information is given in a record's periodical reference field. Subscribers to ISAPOnline have access to complete periodical reference information. The National Library awarded a prize of one year's free subscription for ISAPOnline to Vosloorus High School in Gauteng, for having the highest number of learners at the Careers, Education and Training Faire on 13-14 March 2003.

Support to Other National Library Projects and Functions

The ISAP database is a frequently used source for information searches undertaken for users in the National Library's reference and interlending sections. ISAP is made available to users conducting their own searches on online public access catalogue (OPAC) computers in the Reference Section. Complete articles to ISAP records are provided to SABINET users by the Interlending Section and to non-SABINET users by the reference sections in Pretoria and Cape Town.

ISAP supports the development of a complete periodical collection by identifying missing issues of existing periodicals and requesting substitute copies if necessary through the Periodical Section.

Support is given to the Legal Deposit Section by alerting them to the existence of titles that they have not yet acquired.

With ISAP's contribution, the National Library records South African documentary heritage for posterity.

Generation of Income through Use of ISAP by Other Database Providers and Support of Specific Research

The National Library earns some revenue through the selling of ISAP records or by royalties received. ISAP also indirectly stimulates income derived from information and document delivery functions.

ISAP records made available to SABINET, NISC and the University of Potchefstroom for their licensed use facilitate national and international access to ISAP. Users in the public and private sectors and at universities are reached in this way. ISAP is the most used indexing database on SABINET and the South African Studies CD-ROM published by NISC. The University of Potchefstroom has reported that ISAP stimulates the use of periodicals.

Research in specified subject areas was supported by the sale by special request of selected sets of ISAP records to Taung Agricultural College and the University of Pretoria Law Department.

Selection of Information in Non-scholarly Periodicals illustrated by Examples

Types of Periodicals and Selection of Information

Scholarly journals consist of formal articles accompanied by abstracts and often with keywords. These journals are indexed fully on ISAP and do not pose problems with the selection of information for indexing.

As mentioned above, non-scholarly periodicals (e.g. technical or other subject periodicals) are indexed selectively on ISAP. These periodicals can include much news and product information which are generally regarded as ephemera and excluded from ISAP. However, this kind of information may also include useful facts. These are assessed for inclusion against selection criteria based on practical experience acquired by ISAP.

Criteria for Selection of Information in Non-scholarly Periodicals

ISAP's main aim when it comes to selecting information in non-scholarly periodicals remains to record useful and substantial information applicable to South Africa.

Experience has revealed the following:

- The periodical's subject field or focus should be assessed for inclusion and use by its target market. For example, before including it in ISAP, a periodical focusing on computers but also offering information about human resources should first be weighed against the likely coverage in a periodical that focuses solely on human resources.
- Information presented for a passing reader market (to be omitted from ISAP) should be distinguished from information that could be useful for future reference and should therefore be included on ISAP.

The following can indicate information that is intended merely to attract the attention of readers (as opposed to information that is inherently useful for recording):

- when the information highlighted on the cover page does not even feature on the contents page
- when an item highlighted on the cover page is published as a news item (not as part of a full article)

Process of Selecting Information in Non-scholarly Periodicals

Indexers selecting information from non-scholarly periodicals are expected to consider these relevant points:

- Keep in mind the policy and aim of indexing. It is ISAP policy to omit news items. However, some news items may happen to contain supportive information. This illustrates the problem of vague and indistinct areas in the selection of information for indexing. One option used for handling such information is to refer to a relevant news item in the abstract of a record pertaining to an article – (or usual) – type entry covering a corresponding topic, as shown in Example 1. The news item topic should also be covered in the keyword field of such an entry. Example 1:

TITLE: Servers anchor desktop Linux plans : analysis.
[by/deur] Scannell, E.

ABSTRACT: *Reports from Infoworld(US)* that Linux's overall success is no longer measured by its desktop performance but rather on how it is being used as a mainstream server. Looks at developments taking place in Linux and how its performance in the desktop and server environment is likely to change in the future. **Also see news item "MS/Linux TCO controversy still rages" (p.11)**

IN: *Computing SA* (2003):23:6 p.20 Feb 17
(ISSN) 0254-2196: P11736:

KEYWORDS: SuSE; Linux; Servers; Operating systems; Red Hat

THESAURUS: Computer software; Communication; Business

Another option for ISAP is to make one compound entry comprising more than one corresponding short item, as demonstrated by Example 2. The number of compound entries and information included in compound entries should be limited to avoid unnecessarily cluttering information or obscuring clarity in content. Example 2:

TITLE: Shoden launches modular storage system; EMC Celerra NS600 integrates NAS, high-availability features: **storage & networking: technical**
ABSTRACT: Reports on the following announcements: Shoden Data Systems on the local availability of the Hitachi Freedom Storage Thunder 9500V Series storage systems, with details on specifications and application areas of and technology used in the storage devices. EMC on Celerra NS600 network server combining enterprise NAS operating system with CLARiiON storage architecture for high level information protection and ease of information management

IN: *Computing SA* (2003):23:6 p. 30,31 Feb 17
(ISSN) 0254-2196: P11736:
KEYWORDS: Hitachi Data Systems; Shoden Data Systems; Thunder 9500V; Storage systems; EMC; Celerra NS600; Networked storage; Information protection; Information management.
THESAURUS: Information storage and retrieval; Computer hardware; Technology

- Consult the contents pages for an idea of what the publisher regards as prominent information in the relevant market or industry for that periodical.
- Get an overview of the content as a whole and note how information is presented. This could assist in a process of 'information orientation' in order to select information for a periodical's target group or to fulfil the purpose of the database.
- Assess the usefulness of full reports and articles against selected appropriate short items to which references can be made in the abstracts of records for reports and articles that have been selected.
- Choose meaningful and necessary combinations of more than one short item in single compound entries. Information can be grouped according to form, such as a question-and-answer column, or to subject, where information by more than one author relates to the same or similar topic. Non-related information can also be grouped in compound entries when necessary, as long as the information is clearly identifiable and uncluttered.
- Subdivisions introduced by the publisher can be used in two ways to keep titles in compound entries clearly grouped or identified: 'Storage & networking: technical' is a subdivision found in *Computing SA* (2003).

In Example 2 it is used as a subtitle joining the individual short items.

In Example 3 it is used on its own as a title. In this case each listing of an individual short item in the Abstract is followed by the item's specific page numbering in brackets:

TITLE: Storage & networking: technical
ABSTRACT: Reports on the following announcements: Shoden Data Systems on the local availability of the Hitachi Freedom Storage Thunder 9500V Series storage systems, with details on specifications and application areas of and technology used in the storage devices **(p.30)**. EMC on Celerra NS600 network server combining enterprise NAS operating system with CLARiiON storage architecture for high level information protection and ease of information management **(p.31)**

IN: *Computing SA* (2003):23:6 p. 30,31 Feb 17
(ISSN) 0254-2196: P11736:

KEYWORDS: Hitachi Data Systems; Shoden Data Systems; Thunder 9500V; Storage systems; EMC; Celerra NS600; Networked storage; Information protection; Information management.

THESAURUS: Information storage and retrieval; Computer hardware; Technology

- Evaluate reports by non-South African authors for their relevance to the South African situation to decide how useful it would be to include them. Origin of such information should be acknowledged in the entry's abstract (see Example 1).

Function of the Main Indexing Fields on ISAP

Indexing fields for ISAP are based on the South African Machine Readable Cataloguing (SAMARC) minimum level. Computer programs have been written to convert to MARC 21, to enable ongoing electronic exchange of records.

The subject indexing fields for ISAP comprise keywords in natural language terms, broad thesaurus terms and an indicative abstract. Information in these fields is complementary because in each field the topic of an article can be described on another or further level.

The abstract can be used to indicate the article's focus or intended target group and provide information on how the article is presented, or on the article's relevance to previous or future articles. For further detail see chapter 13 on abstracting.

Natural language terms enable specific description of a topic according to terms used in an article. Additional appropriate terms that are equivalent or related to or descriptive of the article's topic are also allocated here to fully reflect different aspects of a specific topic covered in an article.

The ISAP thesaurus consists mainly of broad disciplinary terms. More than one broad term can be allocated to reflect the cross-disciplinary treatment of a topic.

Such a hybrid system of combining specific natural language keyword terms with broad disciplinary thesaurus terms facilitates a precise recall in searching, as illustrated in the following table.

Precise Recall Made Possible by Specific Keywords Combined with Broad Thesaurus Terms

KEYWORD	+THESAURUS	RESULT
Indexing		26
Indexing	+ Information services	9
Indexing	+ Computers	6
Indexing	+ Medicine	1

Bibliographic description fields include author, article title and periodical reference. The National Library shelf number for the periodical (P number) and its International Standard Serial Number (ISSN) are also made available in the periodical reference field, to facilitate correct periodical identification and expedient provision of complete articles available at the National Library by reference and document delivery services.

Conclusion

ISAP is a dynamic database that continues to develop according to users' needs and technological advancement.

The National Library has systematically worked on improving ISAP according to development plans, identified needs and criticisms or recommendations received. The list below includes examples of improvements.

Periodicals

- The list of 'Approved South African journals, assessed in the reporting year 2005', which is Appendix 4 to revisions of the Department of Education's 'Policy and procedures for measurement of research output of public higher education institutions', has been checked to identify the latest periodicals added to ISAP for more complete coverage of approved scholarly journals.
- In line with reflecting information in a developing democracy, South African published periodicals with a focus on Africa and applicable to situations in South Africa have been added (e.g. *African Finance Journal, African Energy Journal, African Human Rights Law Journal, African Journal of AIDS Research, South African Journal of African Languages, African Studies*).
- Useful and appropriate in-house journals are now included.
- Coverage on music and musicology was extended with titles such as *The South African Music Teacher, South African Journal of Musicology (SAMUS), Scenaria, Musicus* and *Ars Nova*. Other titles complementary to this subject area have also been added (e.g. *South African Journal of Music Therapy, South African Journal of Cultural History, Vuka, South African Theatre Journal (SATJ)*, *Literator*).
- Results of a user survey in 1995 prompted the addition of more subject general periodicals and some popular magazines. Some of these periodicals cover information for developing communities and neo-literates. However, scholarly journals remain ISAP's main focus.

In consultation with the Gauteng Department of Education, periodicals used by school learners and their facilitators, and which could also benefit developing communities and neo-literates, were added to ISAP. A paper copy with records of this subset of educational periodicals dating from January 1997 to August 1998, titled *Edu-ISAP*, was introduced to schools in Gauteng as well as to some public libraries. But this project was discontinued because of lack of interest.

Frequency

- ISAP has been processed and updated on a weekly basis since so requested by SABINET and law librarians in 2002.

Specific Subject Terms

- The ISAP indexing fields described above illustrate the use of a hybrid system of natural language keywords combined with broad disciplinary thesaurus terms. This enables specific searching, as opposed to Library of Congress subject headings (LCSH) previously used and criticised as not specific enough.

Access and Visibility

- Apart from access to ISAP by other database providers referred to previously, the National Library provides Internet access to ISAPOnline at affordable prices in line with different levels of users. Non-subscribers may search ISAPOnline free of charge although access is limited, with incomplete periodical references.
- Promotional presentations on ISAP were delivered at the ISC Schools Library Conference (St Stithians College, Randburg) in 1998, the Community Library & Information Technology Symposium (Boksburg Community Library) in 2001, and at the Association for Southern African Indexers and Bibliographers (ASAIB) Indexing workshop (Unisa) in 2003.
- ISAP participated in exhibitions such as the Educating SA Exhibition at Gallagher Estate (Midrand, Gauteng) in 2000, the Pretoria Show in 2002, and the Careers, Education & Training Faire at Kyalami (Gauteng) in 2003.

Quality Control and Technological Advancement

Towards the end of 2003, the National Library accepted an offer by NISC to sponsor the use of their advanced NISCBase software for ISAP indexing. Training in the new software has started and preparatory work to run ISAP on the new system is under way. It is believed that the use of NISCBase software will greatly improve future content and quality control.

BIBLIOGRAPHY

Behrens, SJ. 1996. Recommendations for the improvement of national bibliographic services in South Africa. *South African Journal of Library and Information Science*, 64(2):81-85.

Bloomfield, M. 2001. Indexing – neglected and poorly understood. *Cataloguing & Classification Quarterly*, 33(1):63-75.

De Wet, LJ. 1955. Compiling the Index to South African Periodicals. *South African Libraries*, 22:93-96.

Ellis, D, Ford, N & Furner, J. 1998. In search of the unknown user: indexing, hypertext and the World Wide Web. *Journal of Documentation*, 54(1):28-47.

Lor, PJ & De Beer, JF. 1994. The State Library in the service of South African science and scholarship. *South African Journal of Science*, 90, August/September:462-466.

Mai, J-E. 2001. Semiotics and indexing: an analysis of the subject indexing process. *Journal of Documentation*, 57(5):591-622.

South Africa. 1998. The National Library of South Africa Act, Act 92 of 1998. Pretoria.

Von Beck, M. 1989. South African bibliographic control tools. *Mousaion*, 7(1):76-86.

Winter, JS. 1967. The Index to South African Periodicals. *South African Libraries*, 34(3):85-89.

9 ADDENDUM: PRINCIPLES OF PERIODICAL INDEXING

Marlene Burger

Abstract

The purpose of this addendum is to provide general information and guidelines for the indexing of continuing resources (serial publications). The focus is on periodicals (journals, magazines). It is a summary of sections from various sources such as Ashcroft and Langdon (1999), Beare (1999), Booth (2001), Lancaster (2003), Taylor (2004) and Wellisch (1995). These sources may also be used for the indexing of other serial publications such as newspapers, trade journals, consumer and special-interest magazines, yearbooks, annual reports, directories, and indexing and abstracting journals.

Introduction

A periodical (journal, magazine) can generally be defined as a continuing publication issued in successive parts with numerical or chronological designations. The medium may be paper, electronic or any other medium. There are two types of periodical index:

- individual indexes to individual periodicals
- broad indexes to a group of periodicals

In the first instance, the publisher of the periodical prepares an index, usually for a volume at the end of a year's run of the periodical. These indexes are prepared under the direction of the editor of the periodical. The approach to this kind of indexing can range from simple, uncontrolled vocabulary to a complex indexing system with a thesaurus.

Indexes of national or even world periodical literature are major enterprises that usually consist of large databases. Such indexes cut across many titles and specific subject areas. ISAP is an example of a national periodical index (see the first part of this chapter for detailed information).

While most books have one or two authors, periodical indexes will list thousands of authors. Periodical indexes are based on the same principles and have the same general objectives as book indexes; but because their scope is broader, they present a number of unique problems. For example, preparing a book index is a well-defined operation, with a beginning and an end. It generally focuses on a general topic, and it can usually be prepared entirely by a single indexer. But periodical indexes are open-ended projects, usually done by a number of indexers, covering perhaps years and with shifts in subject emphases and indexing objectives. Consistency, vital to quality indexing, becomes a challenge. Any one issue of a periodical may deal with a number of unrelated topics by several different

authors, written in different styles and aimed at different users. The periodical index must bring order out of these differences.

With the indexing of periodicals, there is very likely a policy guide that indicates what is to be indexed and what not. Such a policy is necessary in order to maintain consistency when many indexers are involved. The index should include the names of articles' authors and coauthors. Generally, every article should be indexed – these form the core of the periodical. The decision what to index beyond the articles depends on the nature of the periodical, its readers and the economics involved. Some periodicals have only the articles indexed and others may have book reviews, news and gossip columns, and advertisements indexed. The inclusion of such material depends on the degree of pertinent information in them. For example, some editorials contain significant material and often important research analysis, and should therefore be indexed. In general, conference programmes, letters to the editor and the like are not indexed.

A book index is finished and published together with the book (at the back of the book), but with some periodicals the finished index may not be produced until some time after publication of the issues to which it relates. It is also possible for a periodical to be indexed twice, namely currently (either at or close to the date of publication) or retrospectively (to cover material not initially included, or to reflect the needs of a new group of users).

Formats

In this section we concentrate on formats other than printed periodicals.

Ashcroft and Langdon (1999) report an annual rate of increase of about 30 per cent in the number of electronic periodicals, with more than 7 000 titles currently available (please note that this was in 1999). These figures include periodicals on the Internet or CD-ROM or provided by other electronic means. In the academic and research fields in particular, there has been a noticeable increase in the number of periodical titles being published electronically as well as, or instead of, in print. For some titles, the electronic content exactly matches that of the printed version, but for others provision may be of selected articles only, or just a contents list. A smaller number, referred to as e-journals, are available online only. Some titles make the full articles freely available on the Internet, whereas some provide abstracts for free consultation, and others are available only to subscribers or after online registration with the producers.

Publishers may increasingly decide to economise by no longer producing printed indexes and instead directing searches to the online versions of their periodicals, which can be searched in various ways. Indexes may be provided that are similar to the printed versions, but without locators such as page numbers: searchers just click on a heading (author or subject) in the index for the relevant article to be displayed on the screen. Another approach is to use full-text searching by keywords, although using natural language for searching has pitfalls, for example occurrence of words with the same spelling but different meanings (e.g. 'printer' – the equipment or the person?), the existence of more than one word to represent a concept (e.g. dassie, rock-rabbit, Cape hyrax), and variant spellings (e.g. archaeology or archeology). Large subject-orientated databases containing articles from different periodicals can similarly be searched by title word and text word. Indexers

could play a more active role in the design of databases like these, by supplying advice on the most effective methods of providing access for various kinds of user (Booth 2001:198). Users have the same indexing needs, regardless of the entities' physical format. The main issue, however, is quick access from multiple entry points (such as in full-text searching).

Indexes to websites containing the texts of periodicals vary in form. In some, the user has first to select the periodical title from a list of items produced by the same publisher and is then shown a list of year dates to choose from. Selecting a date produces a list of individual issue dates or numbers, and subsequently the contents list of the chosen issue. In others a letter of the alphabet is selected, giving a list of titles starting with that letter; selecting a title gives year dates, issues and contents successively. Author and keyword searches are often possible, either by calling up a list or browsing through it or by entering the name or word in a 'dialogue box' and commanding a search.

Some periodicals, newsletters and bulletins appear in audio form only, or on film and video, designed for users with particular listening and viewing facilities. The index needs of users of these periodicals are the same as those of users consulting printed entities.

For an example of a local electronic periodical index, consult the *Mousaion* index on ISAP.

Indexing Levels

This section is based on Booth (2001:202-207).

In the area of subject indexing, periodical indexing and book indexing differ most. Book indexes aim to represent in fine detail the whole subject content, looking at individual pages, paragraphs and sentences, and including headings containing words that may not appear in the text but that are likely to be sought by users. Indexers of periodicals tend towards general representation or summarisation of each article, perhaps indexing the overall subject and a small number of topics, sometimes referred to, particularly in electronic media, as 'metadata' or 'metatopics'. Indexing of this kind, characterised by less detailed representation of the topics in an article, can resemble the subject indexing (classification) of entities for a library catalogue. Indexing requires great skill in denoting the whole significance of a complete article within a small set of index headings. Some periodical indexes provide no subject index, including instead merely an index of article titles alphabetically arranged according to the first (significant) word of each title. This is of little use to someone searching for subject information, as few users have in mind the exact title of an article when they are searching for information on a topic. Such indexes can complement a proper subject index, but cannot replace it. Another variation on this is to use the article title as a subheading to the chosen keyword or keywords, which is an improvement but still not very helpful.

In the indexing of any kind of entity, the best practice is to provide a heading that represents the topic most directly, so that users can depend on finding what they need at the point of access, not having to consider under which broader heading the topic may be included. For example, an entity on comets should be indexed as 'comets' rather than 'solar system' or 'solar system: comets'. An index consisting of a succession of 'classified' heading sequences, in which sought terms are used merely as subheadings under broader headings, with no possibility of direct access, is unhelpful and should be avoided.

Words from Titles

Sometimes indexing terms are taken directly from the title of the article, but only when the title represents the actual content of the article. Otherwise the quality of the index may be poor because the essence of the articles is not captured and valuable information may be omitted. In many instances titles are poor indicators of content, for example an article about fossilised specimens of dinosaurs, bearing the title 'Locked in rock' would not be helpfully indexed by using the keyword ('rock') from the title. Full titles are sometimes used as a basis for indexes, with the words manipulated automatically – see chapter 10 for information on KWIC and KWOC indexes.

Words from Abstracts

Abstracts usually appear at the head of articles. Before using words from an abstract, the indexer must make sure that the abstract is of good quality and suitable length, and competently written so as to accurately convey the essence of the article. When relying on abstracts as generators of indexing terms, indexers must be careful since the indexing is not done directly from the original articles (unless the indexer is also the abstractor). However, since the indexer is not relying on words from the title only, abstracts provide the opportunity to add as headings any relevant sought terms that are not featuring in the text. See chapter 13 on abstracting.

Words from the Text

Using words from the text provides better representation since the indexer must work through the whole article, reading it attentively, concentrating on sections such as the introduction where the aims and background of the article are described, the methodology used and the outcome (e.g. research results).

Words from Controlled Vocabularies

Using a standard vocabulary such as a thesaurus aimed at a specific subject field provides control, continuity and consistency over long periods of time. There is then no doubt about which term and which spelling of the term should be used. If the thesaurus is updated regularly by adding new topics and inserting suitable cross-references, it will continue to reflect the current and past scope of the periodical. Some thesauri include against each term the date when it was first used, or the bibliographic source from which it was taken. In other instances an obsolete or no longer preferred term carries the date when its recommended status was removed, as well as the term(s) by which it has been replaced. See chapter 14 on thesaurus construction.

Authority Files

If no pre-existing list is used for indexing, it is good practice to build up an authority file of terms used for indexing. This way indexers have a record of terms already in use and

can be sure that they are always used in the same form and that similar headings follow the same pattern. Any names (personal, corporate, geographical) that are included in the index should also be included in the authority file, supported by cross-references from all other forms.

Authors' Keywords

Authors of articles are often asked to provide keywords for their articles to be used in the index. Such keywords may be useful, but they should not be relied on as the sole basis for the index. Authors not conversant with indexing techniques may tend to 'word spot' or provide unsuitable word forms. In the end it is still the indexer's responsibility to ensure that each article is represented adequately with inclusion of relevant cross-references to link related terms, variant spellings and different words for the same topic.

Classification and Coding

Periodical articles can also be indexed by using a system of classification or coding in which topics are represented by alphabetical or numerical codes, or by a combination of both. Well-known classification systems such as the *Dewey decimal classification, Library of Congress classification* or *Universal decimal classification* may be used, or a customised system for the particular periodical, covering its specific subject field and vocabulary. When using such a system, the indexer must identify the indexable topics at the appropriate level in context of user needs, and convert them into the class notations or codes that most suitably represent them (as is done for entities in libraries). Because the notations and codes from the well-known systems are internationally understood, they can be used in indexing and abstracting services to bring together articles in different languages and from a variety of countries.

Guidelines for Indexing

Similar to book indexing, the periodical indexer must be familiar with the indexing process, how to construct entries (for names, scientific vs popular terms, etc.), and how to present entries (this differs from those for a book index and are discussed further on). The following are a few general guidelines:

- Read the title.
- Read the headings in the article.
- Read the abstract/summary that often appears at the head of the article.
- Take a look at the bibliography.
- Scan-read the article.
- Allocate indexing terms (either by using the natural language of the text or by using a controlled vocabulary).
- Take a look at the indexing terms once they have been allocated and decide whether they will guide the user to the article – do they represent the content of the article?
- Work according to the indexing policy of the periodical.

Keep in mind that the existing index (if there is one) to the periodical will indicate what is expected regarding choice of terms and presentation of entries.

It is not possible to discuss or illustrate all the types of periodical index entries an indexer may encounter. The following are some general examples only. It will be worth your while to take a look at a number of periodical indexes to familiarise yourself with the different characteristics.

Examples

The main characteristics of a periodical index are that they may contain certain abbreviations, the locators are made up of different elements, and have author/title information. Presentation differs from index to index, but must be consistent within any one index.

Typical abbreviations used are

A = advertisement

E = editorial

F = feature

R = review

Such abbreviations are not used in all periodical indexes. The abbreviations are considered helpful to the user. For example:

pyramids 61-62(E), 95-105(F), 134(R)

Valley of the Kings 3(A), 25-29(F)

Periodical indexes are presented in different ways, for example subject entries with appropriate titles of articles indented underneath (which may or may not contain the author(s), authors' names with the titles of their articles indented underneath, subject entries with more specific subject entries indented underneath, etc.). It also depends on whether the index is compiled for a single periodical only or for a database containing many periodical titles. In the last instance, the title of the periodical is also included in the index entry.

The following examples of periodical index entries may be useful for identifying the component elements. Note the use of capital letters, italics, bold, et cetera (these may differ from index to index). Each index will have its own rules. These are just examples and are not prescriptive.

Subject Index Entry with Titles of Articles

In this entry the names of authors appear in brackets, followed by the volume number and date of the periodical. Issue numbers and page numbers are excluded. The titles of the articles are not presented alphabetically, but chronologically. This is an entry for an index to a single periodical title:

ARCHAEOLOGY

Digging up the past (Evans, B.C.T.) 1/95

A new era dawns (Smith, B.) 1/95

Ancestors? (Clancy, C.K., Tlhako, E., Botha, W. and Makhanya, M.) 2/96

Egypt's most famous mummy (Hawass, Z.) 4/98

Author Index with Titles of Articles

Here the articles are listed alphabetically under the authors' names:

Botha, W. *see* Clancy, C.K., Tlhako, E., Botha, W. and Makhanya, M.
Clancy, C.K.
    Cradle of humankind 5/99
Clancy, C.K., Tlhako, E., Botha, W. and Makhanya, M.
    Ancestors? 2/96
Evans, B.C.T.
    Digging up the past 1/95
Hawass, Z.
    Egypt's most famous mummy 4/98
    Nefertiti 5/99
    Tomb robbers 4/98
Makhanya, M. *see* Clancy, C.K., Tlhako, E., Botha, W. and Makhanya, M.
Smith, B.
    A new era dawns 1/95
Tlhako, E. *see* Clancy, C.K., Tlhako, E., Botha, W. and Makhanya, M.

In the example of Clancy and three other authors, one may also use 'and others' or '*et al.*' in the cross-references: Botha, W. *see* Clancy, C.K. and others 2/96, or Botha, W. *see* Clancy, C.K. *et al.* 2/96.

Cumulative Index

The locators in a cumulative index to a single periodical (covering, say, five years) must include sufficient date and volume information to guide the user to the correct issue. The following index entries are alphabetically arranged according to title and include year, volume, issue and page numbers. Authors are excluded:

Ancestors? 1996 2(3):34-40
Clay tablets and their hidden messages 1997 3(1):11-18
Cradle of humankind 1999 5(3):52-62
Digging up the past 1995 1(1):50-58
Egypt's most famous mummy 1998 4(4):22-31
From clay to papyrus 1996 2(3):15-18
Museums as social institutions 1997 7(1):43-47
Nefertiti 1999 5(4):71-75
A new era dawns 1995 1(6):10-18
Tomb robbers 1998 4(2):42-45

Index Covering Several Periodical Titles

In an index covering several periodical titles, full details of the containing periodical will be indicated, including title, volume number, issue number, page numbers. Authors are excluded and may appear in a separate author index. For example:

Ancestors? 1996 *Archaeological Studies* 2(3):34-40

Dinosaur finds. 2001 *Journal of Palaeontology* 4(4):29-35

Evolution vs creation. 2000 *Bulletin of the SA Archaeological Society* 3(1):69

All the examples in this section illustrate that periodical indexes differ to quite an extent. It cannot be stressed enough that the indexer must work according to the rules (policy) for a particular index.

Teamwork

As with most cumulative indexes, indexing is done on a cooperative basis – employing more than one indexer (mostly freelance indexers). This implies that each member of the team must understand and observe the rules of the indexing policy. There must also be the opportunity to exchange information and discuss problems. See the section under 'Contracts' regarding ISAP.

BIBLIOGRAPHY

Ashcroft, L & Langdon, C. 1999. The case for electronic journals. *Library Association Record,* 101:706-707.

Beare, G. 1999. *Indexing newspapers, magazines and other periodicals.* Sheffield: Society of Indexers. (Occasional papers on indexing; no. 4).

Booth, PF. 2001. *Indexing: the manual of good practice.* München: Saur.

Lancaster, FW. 2003. *Indexing and abstracting in theory and practice.* 3rd ed. Champaign: University of Illinois, Graduate school of Library and Information Science.

Storage & networking: technical. 2003. *Computing SA,* 23(6), 17 February.

Taylor, AG. 2004. *The organization of information.* 2nd ed. Westport, Conn: Libraries Unlimited.

Wellisch, H. 1995. *Indexing from A to Z.* 2nd ed. New York: Wilson:347-354.

10 AUTOMATIC INDEXING

Peter Underwood

Abstract

Using a computer to process a text that is already in machine-readable form so that index terms are allocated to its content without direct human intervention can be described as 'automatic indexing'. There is, however, a significant gap between the systems currently available and this objective. The author explores this in some detail but concludes that it is a development of the future and that the human element is still a vital component in indexing.

'Drowning in information but thirsting for knowledge' represents the paradox of our times. The development of Information and Communication Technologies (ICTs) has changed the way in which many aspects of the 'information game' are played but has not affected its fundamental principles. The problem those principles seek to solve was articulated by Plato in the *Meno* (Stanza 80e): '[A] man may neither seek what he knows or what he does not know, because he does not have to seek what he knows, for knowing it, he need not enquire, nor may he seek what he does not know for then, not knowing it, he knows not what he is seeking.'[1] Despite significant progress in the last 20 years in developing computer systems that can rapidly retrieve text as sentences, words and word stems, the essence remains a gamble between author and potential reader. If both use the same word to describe the same concept, then the retrieved document *may* satisfy the reader.

Note the emphasis on 'may': the problem is more subtle than simple matching of single words or, even, sets of words. The capacity of a language to encapsulate fine distinctions means that even texts with a high degree of overlap between words used by the author and those used by the searcher can still fail to capture what it is the searcher is looking for. How we impute and recognise meaning is the subject of investigation by psycholinguists and it is clear that the development of strong theories and robust techniques is still some years away. Unless there is a significant breakthrough in understanding and modelling the complexity of textual communication and the association of meanings, it will not be for this generation to stand, as they might on the deck of the Starship 'Enterprise', and command the computer to perform a complex analysis resulting in retrieval of only a few highly relevant documents.

1 I am grateful to Jennie Underwood for this translation and for first drawing my attention to the significance of Plato's comment.

Yet this is a reasonable aim. Indexing by human labour is expensive and labour-intensive. It is also notoriously difficult for humans to agree on 'aboutness' and what words should be used to describe it (Blair 1990). Even more difficult is to ensure that the words used by searchers correspond to those used by indexers (Lancaster 1991:10-14).

Mechanising index production would clearly have benefits in speeding up production, replacing human with machine time – and thus reducing costs of production – and ensuring greater consistency. However, the present limitations of our understanding of psycholinguistics still means that the rules by which computers produce such indexes are unlikely to result in indexes where there is a perfect match between what is represented and what is sought. Lancaster has suggested that it may be more fruitful to concentrate on assisting searching rather than seeking greater consistency of indexing (Lancaster 1991:74-75).

Using a computer to process a text that is already in machine-readable form so that index terms are allocated to its content without direct human intervention can be described as 'automatic indexing'. There is, however, a significant gap between the systems currently available and this objective. 'Allocation to content' requires an understanding of a text, imagination and a mastery of language in order that the meaning of the content is expressed. A commonly used approach has been to use the computer to create concordances, or lists of words, which the searcher then explores.

'Word association' prompts the searcher to link target words together, using logical (sometimes called 'Boolean'[2]) operators, into a search statement. The role of the computer is firstly to scan texts and create an 'inverted file' which associates each word in the file with positions in the texts. Its second use is to match each word in the search statement against the inverted file and to identify texts that have words in common. These can then be retrieved and scanned for relevance. By suppressing very common words with low subject significance, such as 'and', the inverted file is kept within manageable proportions. Because the searchable word list reflects only words in the texts, the technique may be described as 'derived' or 'extractive' indexing.

Computers are excellent tools for creating and managing such lists, so that the addition of a document to the database results in the automatic inclusion in the inverted file of the words it contains. Computers are also highly suitable for searching the inverted file, matching the pattern of a sought word against the patterns contained in the file.

Automatic indexing is the creation of index terms from a text or image by computer, without intervention by a human indexer. With the widespread use of computers to store text, a potential for confusion has arisen regarding the indexing function. 'Text searching' is the scanning of a text to match terms decided by the searcher; it is sometimes called 'natural language searching' to contrast it with 'searching by index' where the search for terms is conducted in a file separate from the text and the text and index are linked by a locator such as a page number. The text can be a complete work (often called 'full text'), or some representation of it, such as an abstract or title. Many electronic sources of information,

2 Named after George Boole (1815-1864), Professor of Mathematics at Queen's College, Cork. The diagrams sometimes used to depict logical combinations, 'AND', 'OR' and 'NOT' were development of the English logician, John Venn (1834-1923).

held in databases or accessible over the Internet or World Wide Web, are available for full text search. The low cost of computer storage and the speed and power of the processors have meant that this approach is both feasible and economical to provide.

Maintaining the distinction between indexing and text searching is crucial if one wishes to consider effectiveness, however. The assumption that text searching can invariably offer 'better' results than searching based upon indexes arises because it is the whole representation of the text that is available for searching rather than terms considered, through some process of selection, to be significant indicators of content. The argument is based on the presumption that searchers can best describe and recognise what will be significant. It also begs the question 'What is "better"?' Once a text has been transferred to a form suitable for computer storage, a text searching approach to location of relevant items shifts most of the costs of searching and shifts *all* of the effort required to the searcher. This is certainly better for the provider of a source of information and may even give the illusion to the searcher that this is a better product – but it is not necessarily borne out by experience.

Since the 1950s, there has been a considerable volume of research on the effectiveness of methods of indexing and comparison with other approaches such as text searching. In addition, practical experience in using databases and searching the World Wide Web and other Internet channels has yielded contradictory, and sometimes puzzling, results. It cannot be safely concluded, for example, that either approach is, of itself, superior even for the fundamental task of all information retrieval, that of indicating to the searcher highly relevant documents in response to a search.

The fundamental task is, however, only one aspect of how 'effectiveness' might be measured and assessed. A full assessment must take into account the costs of input, costs of storage and production and costs of searching, and compare these with the effectiveness of the search. Assessing effectiveness is also fraught with difficulties: the aim of searching is to retrieve *relevant* documents and a minimum of irrelevant documents; but, surprisingly, even experts in a subject area are not consistent in their judgments of what constitutes a document that is an acceptable answer to a search (Cleveland & Cleveland 1990:147). 'Relevance' is also a complex concept. At one level, one may judge the relevance of a document in terms of how well the search terms used to retrieve it match the contents of the document; on another, there is a judgment on how useful the document may be to the searcher. This aspect of 'pertinence' is obviously much more subjective and temporal than relevance per se, and takes into account the effort that may have to be expended on locating the document, as well as its intrinsic worth in solving the searcher's problem. The speed of a computer search may provide the illusion that the results are also going to be speedily assessed. Practical experience of searching on the World Wide Web should disabuse one of this assumption, however: a search retrieving several thousand apparent 'hits' is hardly effective if it takes several hours to assess the relevance of the items found, even assuming that the searcher is persistent enough to evaluate the whole set.

Text searching approaches to information retrieval depend on documents being available and accessible on a computer system. The provision of information in the form of databases is the most prominent example and it is common to find full text searching facilities provided as a means of information retrieval. It is also quite common to find that

the records in such databases are enriched with subject descriptors assigned by an indexer, or added automatically, from an index vocabulary. Research on index languages has explored the utility of searching using assigned index term systems, compared with natural language text searching. The results of early studies suggested that the assignment of index terms did not convey any advantage over natural language approaches and this finding was confirmed by further experiments. Later work has resulted in some modification of this view, especially when large databases and full-text systems have been considered: the problem of overwhelming the user with retrieved documents is recognised as being significant. Because of the richness of natural language, it is highly probable that searches of full-text systems will retrieve many documents but will also retrieve many that are irrelevant, or only partly relevant.

Despite this consideration of effectiveness, the cost of manually assigned indexing may eventually serve as a deterrent to their production, especially when sources such as electronic books and journals become more common. 'Bundling' and electronic text with software capable of natural language searching of the full text is likely to be a cheap option, attractive to the publisher because it is fast to produce and update and superficially attractive to the purchaser because it appears to offer unlimited possibilities of searching. The 'value-added' component that can be supplied by a skilled indexer may be relegated or diminished in importance *unless* publishers and purchasers can be persuaded of the virtue and power of assigned indexing, even as an adjunct to natural language searching. The desirable product is a text that can be searched using both means: index terms assigned by an indexer after inspection of the document combined with natural language searching ability, when required.

The starting point for any process of automatic indexing is input of the text to the computer. The text is allocated a unique document identifier by which a copy of it can be subsequently located in computer storage. The characters of the text are then analysed to create a list of lexemes, or word-like units such as chemical formulae, that it contains. Each lexeme or unit is linked back to its parent document using the document identifier, this creating an 'inverted file'. Lexemes derived, or extracted, from the text provide the raw material on which automatic indexes are produced.

The earliest applications of this approach were the indexes compiled by Hans Peter Luhn (1896-1964) and Herbert Marvin Ohlman in November 1958, when the International Conference on Scientific Information (ICSI) took place in Washington, DC. Luhn produced indexes for the conference programme using his 'Keyword in Context', or 'KWIC' technique. Ohlman demonstrated a similar approach, which he called 'Permuterm' indexing, and used it to produce an index to the proceedings of the conference. The impact of the 'information explosion' following from the upsurge in economic, technological and scientific activity after World War II was the driving force behind efforts at improving documentary control. Speedy production of indexes and searching aids was considered essential if the growing volume of scientific material was to be made quickly accessible. In addition, by applying simple rules the lack of consistency implicit in human indexing would be avoided. Luhn (1957), Ohlman (1957) and Phyllis B. Baxendale (1953), who developed techniques that identified topic sentences in documents, were pioneers in the field of applying computers to the task of mechanising the production of indexes, though many others contributed to

the search for techniques that could contain the 'information explosion' and aid searchers to find relevant documents. The KWIC index based on the titles of documents is the most familiar product of automatic indexing from this period.

At the core of these approaches is the concept of deriving keywords from some part of the text of a document by machine manipulation. Statistical analysis identifies very commonly occurring words, such as prepositions, coordinators and articles that are low in subject content, but needed for syntax; these are allocated to a 'stop list' that excludes them from indexing. Such approaches may be described as 'concordance-like' and, although they can be rapidly produced, they do not provide means of searching other than through the words used by the authors of the documents processed for the index. In subject domains where the vocabulary is more limited and standardised than might be encountered in, for example social sciences, the degree of consonance between searcher and author may be high, thus obviating the disadvantages of the purely derived approach.

The derived index approach may be summarised as a quick and economic approach to index generation with no human intervention, where the costs and effort have been shifted from the creation step to the searching step of the index production and use. It is the searcher who has to display some flexibility of mind to think of the range of terms that authors may have used to describe a sought phenomenon. Despite this, and taking account of the costs of human intervention in indexing, Lancaster has suggested that it may be more fruitful to concentrate on assisting searching than seek greater consistency of indexing (Lancaster 1991:74-75). A familiar aspect of this problem is that of seeking relevant documents on the World Wide Web using a search engine.

Several strategies have been developed as a means of assisting searchers. Truncation (also known as 'stemming') allows for searching on a portion of a word; 'leftmost' truncation is commonly used, suppressing the word ending so that singulars and plurals of words that follow the standard format can both be matched. 'Rightmost' truncation allows for the suppression of prefixes, while internal truncation, or 'wildcards', allows for suppression or substitution of letters within words. These latter forms of truncation are often useful for specialised forms of searching, such as for molecular formulae. In each case, the result of truncation is to increase the number of matching documents during a search; however, this will almost certainly also increase the number of irrelevant documents also retrieved. The more extensive the truncation, the more documents will have to be scanned by the searcher and the greater the likelihood that many will have to be rejected as irrelevant. Weak truncation tends to be the most effective because it balances the ability to suppress usually insignificant differences such as singular and plural word forms, with retention of enough of the word stem to provide discrimination.

The list of words in the document can be further processed to analyse their distribution. By noting the frequency with which each word appears, some judgment can be exercised on whether it is likely to be significant as a descriptor of document content. Words or terms appearing infrequently are unlikely to be good descriptors of content, whereas those appearing with high frequency may not be good discriminators. Candidate words or terms can then be chosen from within a suitable frequency range. Such a list is, at least, an aid and starting point for an indexer but is still likely to include many words that are syntactic rather than descriptive. Automatic indexing systems need to be able to

recognise word forms and functions to highlight those that are most likely to connote subjects. Reliance on word frequency is also not effective once many texts have been added to the database: what becomes important is the discriminating power of a word or term within the database – that is, across the corpus of documents that it contains – rather than within a document. The discriminatory power of a word or term is also sensitive to context: what may be a common word or word form in one type of document will be uncommon in another. The discriminatory power of a word may be improved by assigning it a weight, such as being a major or minor descriptor – but this again raises the question how to recognise the appropriate thresholds to make such a choice. Simple reliance on computer analysis of word frequency is thus unlikely to produce indexes that are effective for searching.

Syntactic analysis has been explored as a step beyond word frequency approaches. It is an attempt to recognise the form and function of a word, or group of words, in a phrase. By 'parsing' a sentence according to rules that define parts of speech, the computer processing yields a series of candidate phrases that are deemed to describe it. This is presently an expensive exercise both to develop and to implement, and there is a lack of evidence that the information retrieval is sufficiently improved to justify the attendant costs. It remains an important area for further research.

Automatic indexing applied on a large scale must be regarded as a development of the future, allied with greater computer power and a much deeper understanding of how human beings recognise, weigh and assign meaning to ideas and documents. For the present, computers can provide a useful function by rapid production of simple extractive indexes and natural language searching. Human indexers still have an important role to play in providing indexes to conventional publications and, arguably, to enriching the retrieval potential from databases and other electronic sources of information.

BIBLIOGRAPHY

Baxendale, PB. 1953. Machine-made index for technical literature: an experiment. *IBM Journal of Research and Development*, 2(9):354-361.

Blair, DC. 1990. *Language and representation in information retrieval*. New York: Elsevier.

Cleveland, DB & Cleveland, AD. 1990. *Introduction to indexing and abstracting*. 2nd ed. Englewood, Colo: Libraries Unlimited.

Lancaster, FW. 1991. *Indexing and abstracting in theory and practice*. London: Library Association.

Luhn, HP. 1957. A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, 1(4):309-317.

Ohlman, H. 1957. *Permutation indexing: multiple-entry listing by electronic accounting machines*. Lexicon, Mass: Rand Corporation, System Development Division.

11 WEB INDEXING

Madely du Preez

Abstract

The meaning of web indexing (or online indexing) is explained, with advice on how to index a website. Some web indexing decisions and questions are explored, focusing on indexing a single website for navigational and retrieval purposes. A brief explanation of HTML (HyperText Markup Language) is given to introduce meta-tags as a means of controlling how search engines index a website. This is followed by a brief discussion on the description and keyword meta-tags and their creation.

Introduction

Indexing, as in 'indexing the Internet', is an approach to information discovery in a way that lends itself to effective navigation across distributed repositories (Denenberg 1996:1). *Information discovery* means locating objects of interest when the population of objects from which to choose is potentially widely distributed. Objects may be aggregated into collections, and organised thematically, but may also be distributed across the Internet – even though the collection appears logically cohesive because the metadata structure associated with the collection is supporting coherent navigation.

The term 'web indexing', or 'online indexing', covers a wide range of processes and products, from putting a book index online to attempting to create an organised index of the entire Web. The most common web indexes could include static indexes, online publications (e.g. online newsletters), combination print-online indexes, website indexes, site maps, web directories, selective web indexes, indexes of web indexes, keyword indexes and controlled vocabularies/thesauri. Usually, though, web indexing simply involves the creation of hyperlinks within documents that are accessible on the Web (Maislin 2003:3).

Denenberg (1996:1) distinguishes between two broad web-indexing models: a *global* index versus independent *local* indexes. Global indexes are created by search engines like Yahoo and Google or indexes for an intranet where each department, section or branch of the organisation to which the intranet belongs is responsible for its own website. Since the current model of the Web is a single, logical agglomeration of documents, with logical flat indexes of its entire content, global indexes are also physically distributed. A local index is an index created for a specific website, such as the departmental or branch website mentioned above. This index will not be distributed across the Internet in the same way as global indexes.

Website indexing (Rowland 2000b:1) involves two basic skills: indexing and website design as well as the willingness and resourcefulness to invent new ways to present information. Browne and Jermey (2001:47) add an understanding of the structure of information on the Web to these skills. The best form of a website index is highly dependent on the type of website being indexed and the needs of the audience for whom the index is being created (Rowland 2000a:1). As a website index aims to improve access to information on the website, the success of the index can be measured by how quickly users can get in and out of the index page and find the desired information (Maislin 2000:40).

The purpose of this chapter is to explain what web indexing (or online indexing) is and advise on how to index a website. Some web indexing decisions and questions will be examined. The focus will be on indexing a single website for navigational and retrieval purposes. A brief explanation of HTML (HyperText Markup Language) will be given to introduce meta-tags as a way of controlling how search engines index a website, followed by a brief discussion on description and keyword meta-tags. This will include an example on the creation of meta-tags.

Web Indexes

Indexing the Web involves three different kinds of indexing: a book style of hard-coded index links within a website to serve as a navigational tool, subject trees of reviewed sites, and indexes created by search engines (Denenberg 1996:1). The most important characteristic that indexes on websites and the Internet share with print indexes is the goal of any good index: 'to bring together all the items on a similar subject which are separated in the book [or website] itself' (Broccoli & Van Ravenswaay 1999:37). The scope, structure and design of a web index is however different from traditional indexes. Cleveland and Cleveland (2001:220) explain that a book index usually points to information in one file (the text of the book) and periodical indexes point to information in several thousands of files (numerous journals). Web indexes however point to many millions of files, at many levels, existing in every niche and corner of the world.

Web indexing could, however, also refer to the use of meta-tags where keywords are assigned to a web document to assist search engines in correctly indexing the document in their indexes.

The term 'index' has developed a broader, more general meaning on the Web and could essentially be a collection of Uniform Resource Locators (URLs) designed to take users to other websites that have either been collected manually or by search engines like Yahoo. Such indexes usually reflect little analysis of the websites they list and these lists are more like catalogues than traditional indexes.

All indexes have page numbers, unlike the World Wide Web. Embedded indexing techniques, however, do not require the indexer's awareness of page numbers (Maislin 2000:43), which makes it an ideal technique to index web documents. The indexer merely creates hyperlinks to the nearest section titles (or the specific point in the document where the term occurs) instead of page numbers to aid navigation among documents (or within a specific document) that are accessible on the Web (Maislin 2000:45).

Indexing Pitfalls

Website indexes can enhance search success, but are costly to implement and, depending on the experience of the indexer and the time available for indexing, are of variable quality. There will inevitably be a time delay between the creation of the web page and the addition of links to it in a search engine's index. These delays can be reduced if the page is reviewed for indexing before being loaded to the Web (Browne & Jermey 2001:28-29).

It is important to bear in mind that online or Internet users might find it fun but also diverting to surf the Web. Those users who make use of indexes are generally looking for specific information, and it is the indexer's task to help them find it. Lathrop (2003:3-4) cautions that readers are more easily frustrated when searching an online index because they do not have the advantage of spatial orientation, which they are familiar with in printed books. The sheer size of the online document may cause them to think scanning the document for the information they want would simply be an exercise in futility. A well-written, comprehensive index increases customer satisfaction and reduces costly product support time because it makes a product easier to use.

Web Indexing Decisions and Questions

A number of indexing decisions and questions need to be made and asked at the beginning of a web indexing project.

What Are the Boundaries?

Will information be added to this website, like a corporation's intranet where new memos, presentations and budget reports are added to the site which would constantly change the indexing? (Wright 1997:3). If so, it could be advisable to create a thesaurus or a terminology-orientated ontology to assist in building a structured index of the website (Desmontils & Jacquin 2003:1). In this context, an ontology is a 'set of concepts each one represented by a term (a label) and a set of synonyms of this term, and a set of relationships connecting these concepts by the specific/generic relationship' (Desmontils & Jacquin 2003:[3]).

Web Authoring Tools

These are software packages that build raw information into usable online resources. These tools operate in several ways and one would have to determine the orientation of the project (Wright 1997:4).

How Will the Index be Displayed?

It is important to know how the index will appear to the user – how many levels of index will appear? When the user clicks on an entry will a list of topics relative to the applied term appear, or is there only a one-to-one correspondence between index entry and topic? Are there any controls over the sorting of the entries? How will cross-references be put in, and what will happen if the user clicks on them? How will the user get to different sections of the index: by clicking a button or typing a letter? (Wright 1997:6-7).

How Does the Indexing Get Into the Files?

Closed systems force the indexer to work within a closed software package, where only that tool can be used to input entries into the system. This would usually require the indexer to operate at the client's worksite so that their software and files can be accessed. Open authoring tools will however allow the indexer to import and export indexing into and out of its file structures, usually with spreadsheet-like data files (Wright 1997:7).

What Kinds of Files are Included in the Project?

Online indexes can lead a user to a topic, to photographs or sounds, a menu of choices the user needs to make, to a website, or the index connection can start a macro running, opening files or formatting text. Wright (1997:8) advises one to ask about the structure of the project and its menu screens if they are used. She explains that menu screens are like table-of-contents screens that present a user with choices, which then lead to specified topics. Another type of topic structure one needs to ascertain concerns decision trees and whether any further help screens are going to be offered.

Who is Inserting the Indexing and Maintaining It?

Who will be maintaining the indexing efforts, and how will that process work? Expanding indexes usually require a master indexing list of used terms to be maintained, or a complete thesaurus (Wright 1997:9).

Time Frames

- Various factors, such as authoring tool problems, debugging glitches and editing difficulties, could complicate the completion time for a web indexing project – working with a known tool that has been used before will reduce the work schedule while unknown and new tools need more time for testing and compiling (Wright 1997:9).

Equipment Needs

It is advisable to determine the necessary kinds of equipment before beginning the project, especially when working off-site. It is very important to ensure that the computer involved can handle the tools to avoid memory or incompatibility problems. Basic equipment needs include (Wright 1997:10)

- a computer with enough memory and storage
- modem or network card
- authorisation and passwords to access files
- ability to FTP or have remote access if off-site
- authoring tool(s) installed correctly with all subsidiary software and utilities
- capability to run authoring tools without conflict

Availability of Files

When working in-house it is important that one has access to the authoring tools and the files to ensure sufficient time to complete the project. When working in stand-alone indexing software, and then embedding or inserting the indexing into the authoring tool, it is important to get print-outs or copies of the files to work from. The indexer should also consider how files will be transported (Wright 1997:11).

Trial-run Indexing

Consider doing a few sample entries, and then do a trial run of the compiled index to ensure correct and smooth running. Avoid entries that could cause the computer to hiccup: sorting problems, special characters, odd punctuation, cross-references, special jumps to animations or odd topics, things that might not work as planned (Wright 1997:11).

Fees for Online Projects

Wright (1997:11) advises charging an hourly rate for web indexing projects unless one is working for a group that has established its procedures and used the authoring tools several times before.

Indexing as a Navigational Tool

One needs to consider two indexing processes when indexing a website or a web document. The first is indexing as a navigation tool where the purpose of the index is to transport the user between various points of interest within a website or number of websites. The second process is to assign keywords or indexing terms using meta-tags to control how search engines index the website for retrieval purposes.

An entry in a book index contains much information. Consider the following single entry:

Multiple indexes, 49, **66-67**, 137

Users know that the bulk of the information on multiple indexes start on page 66 and can assume that page 49 provides an introductory comment, while page 137 casually mentions multiple indexes since it comes completely isolated from the other page numbers.

An online or web entry would look something like this:

Multiple indexes, *,*,*

The user no longer knows which entry is the most important and can no longer understand how these documents fit into the entire collection of documents. Neither is it possible to know the lengths of the documents themselves or to determine the order of the entries. Maislin (2000:44) explains that this loss of page numbers and of global context presents the indexer with a fundamental handicap to writing a good index.

Most of the other dilemmas facing indexers are related to layout (Maislin 2000:44). In the main authoring software designers pay little attention to the need for indexes or the indexing process, since these are often merely added as fancy enhancements. Some do not allow for sub-sub-entries, cross-references or special types of references, nor do they allow for locator-specific formatting. HTML (HyperText Markup Language) also does not allow for indentation.

The following sections will examine some ways of accomplishing these goals and overcoming some of the online environment challenges.

Rating Index Entries for Importance or Relevance

There is no intrinsic way to present ranges of information on a website. Large quantities of information are also split up due to *document chunking* (the breaking up of documents into small units). As with books, indexers can write locator text that explicitly communicates context by making use of a section title or a heading to provide the locator. The locator text is however the web indexer's only opportunity to provide context. It is advisable rather to use descriptive headings so that the user can confidently guess what these sections contain. For example (hyperlinks used instead of page numbers and represented by underscoring):

Maintenance, in web indexing
    Markup language
    Mediation of discussing list disputes
    <META> tags
    Multiple indexes

Maislin (2000:45) is of the opinion that the best way to improve usability is to avoid ambiguity. Even one-word locators like 'Marketing' and 'Networks' are unambiguous to most users.

Visual cues, such as colour and enhanced font styles, can be added to online or web indexes to make the index more meaningful. More important entries can be presented with larger lettering while links to definitions can be coloured red. However, these should be used sparingly and only for clearly defined purposes. Note that an index with a cornucopia of line sizes and colours looks unprofessional.

Book indexes are arranged alphabetically. Maislin (2000:46) suggests using a sort order that makes more sense to the user in an online or web index. Possible sorting schemes include *chronology*, *importance* and *page order* (for printed books converted to online documents).

The online or web indexer can also rate entries by displaying those that meet a pre-defined threshold of importance in bold – a binary rating system. But this system is not as effective when applied together with subjective thresholds like 'importance' or 'relevance', because the user does not necessarily understand that threshold. The only advantage of a binary rating system is that it is easy to implement.

A discreet rating system with three to five levels is better, the most intuitive being the 'asterisk rating system' where the additional * represents slightly greater importance or relevance.

Communicating Relationships between Entries

Cross-references are more complicated online or on the Web, but are essential in a stand-alone index since they communicate the relationships between index entries. *See* references direct users from terms that are not used to terms that are, for example 'Unisa, *see* University of South Africa'. *See also* references are used with terms that do have links, to suggest additional search possibilities. *See also* references are usually reciprocal, and one would expect to find the reference '*Haemophilus influenzae, see also* Meningitis' as well as 'Meningitis *see also Haemophilus influenzae*'. Note references that contain a scientific name have to be in italics next to an italic *see* or *see also*. When both terms in a reference are in italics, the *see* or *see also* is written in roman font (Browne & Jermey 2001:41).

Maislin (2000:49) gives the following tips on creating cross-references:

- *See* cross-references should be used carefully because of the difficulty users have with navigating long online indexes. Rather post index information at both locations if two main-level terms are effectively identical.
- *See* references can become a technique to manage users' word choices and educate them towards the preferred vocabulary set. This can be very important if the index is enhanced by a search engine, especially as search engines do not handle synonyms or related terminology.
- *See also* cross-references should be presented as early as possible, like at the top of a sub-entry list as opposed to the bottom. This should help users understand the document structure.
- *See also* references could also be used to point to other locations within the index, appropriate reference lists within the document, transitory web pages and websites outside the documentation.
- Use circular *see also* references in a hyperlinked text so that users can retrace their steps.
- 'Related Topics' paragraphs in the documentation work similarly to *see also* cross-references.

An interesting alternative to multiple cross-references is the creation of multiple indexes. This approach can make a website more comprehensive without overloading it with text. For example:

Engineering Institutes in South Africa, by discipline:
Acoustical | Aeronautics | Chemical
Civil | Electrical | Electronic
Industrial | Mechanical | Mining | Structural
Clicking on Electrical will display the second index with the names of the different Electrical Engineering Institutes in South Africa.

The creation of multiple hyperlinks for a single locator remains the most obvious technique to demonstrate relationships between index entries: 'Multiple indexes, 49, 66-67, 137' could now be presented using a multi-lined locator approach or simple numbers when images are presented: 'Pictures of Table Mountain, 1, 2, 3'.

The following is an example of an HTML index. Note that this layout is reproducible using standard HTML definition lists. Entries are sorted either alphabetically or by importance, depending on the context. Qualifiers could be added while most sub-entries get their own line. The lists of cross-references are combined horizontally, with semicolons between them:

Professional organisations in South Africa
Association for Consulting Engineers
Council for Professional Engineers
See also Engineering Institutes in South Africa, by discipline: Acoustical; Aeronautics; Chemical; Civil; Electrical; Electronic; Industrial; Mechanical; Mining; Structural
Information Science interest groups
See also Association of Southern African Indexers and Bibliographers; ASAIB; Interest Group for Bibliographic Standards; IGBIS; LIASA interest groups; Research and Training Interest Group; RETIG; South African Online Users' Group; SAOUG
South African Institute of Architects
South African Institute for Library and Information Science (SAILIS): *See* Library & Information Science Association of South Africa; LIASA
South African Teachers' Association

When choosing a design for an index, one should always ensure that enough space is allowed for qualified links, enough vertical space for entries and sub-entries, a visible white space between topics and subtopics, clear and unambiguous lists of cross-references, and no conceptually overlapping sub-entries (Maislin 2000:51).

Developing a Visual Architecture and Navigation System

Rowland (2000b:1) finds it necessary for one to know basic HTML commands before starting to index or at least have an understanding of what HTML is and how it works, as well as know how to construct an index in plain HTML. This will help to format the index page more precisely and it will also help identify and solve HTML problems. (HTML will be discussed later.) Maislin (2000:51) reckons it is not uncommon for an author or indexer to spend more time designing and editing the format than actually writing the index or adding context to the locators.

One way to solve the problem of having a single web page with thousands of lines, where a user has to scroll down to the second last entry in the index, is to divide the index into various pages. The most common solution is to break such a long index into pages by letter. A navigation bar of letters, with each letter linked to one of the index pages is placed at the top of each page:

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

Letters could also be grouped together or divided further. It is however very important to be consistent. The navigation bar could also be repeated at the bottom of each index page, to enable the reader who wants to jump ahead to do so without having to scroll back to the top of the page. The navigation bar (along with the Next and Previous buttons) could also be placed in a frame that does not scroll, at the top, bottom or side of the page (Maislin 2000:52).

Indexing for Internet Retrieval

The World Wide Web lacks a systematic structure by which an information seeker can find desired information in the minimum time. Broccoli (2003:1-2) explains that many search engines do free-text searches. This means that the engine is looking for every occurrence of a word within the text of a page and a document could thus be retrieved even if the word used in the search appears once in the whole document. The solution lies in assigning keywords and setting search engines to search only keywords, developing a custom-designed thesaurus based on the terminology used on the website as well as within the specific industry or profession, and setting the search engine to recognise synonymous terms.

The assigned keywords are then entered into the header of the website using meta-tags. Meta-tags are thus a means to control the process of web indexing (Alimohammadi 2003:238).

What are Meta-tags?

Meta-tags are non-displaying, or hidden, HTML tags that may provide site owners and authors with a degree of control over how a web page is indexed. The term 'meta-tag' actually arises from the term 'metadata' which means data about data that is structured to describe an information object or resource. As such it may describe the content of a web page (Alimohammadi 2003:238). Meta-tags have become a design element used by webmasters and indexers to help search engines to better acknowledge information-poor pages (Sullivan 1998:3). Website designers or indexers use them in the indexing process of a website by choosing keywords describing the content of the site. They also provide a mechanism for identifying information that should be included in the response headers for an HTTP (Hypertext Transfer Protocol) request and can be extracted by servers and clients for use in identifying, indexing and cataloguing documents.

The Concept of HTML

HTML, the computer language used to create web documents, is written using tags, which are commands written inside angle brackets: <and>. Rowland (2000b:2) explains that all tags are written in pairs, with opening and closing tags surrounding the affected text. The opening and closing tags are identical except that the closing tag contains a forward slash symbol (/). In <B>Association of Southern African Indexers and Bibliographers</B>, <B> is the opening tag and </B> the closing tag. When the document is viewed in a browser,

one would not see the tags, but will see the formatting: 'Association of Southern African Indexers and Bibliographers'.

It does not matter whether the 'B' is typed in the upper or lower case. HTML editors might automatically change all tags to upper case. It is therefore best to be flexible.

All web pages have two sections: a HEAD section and a BODY section. The information about the web page goes into the HEAD section, including the title and any information one would use to describe the web page with, such as the information included in meta-tags, for example the keywords.

The title for the ASAIB website will appear as follows in the HEAD section:

```
<HEAD>
<TITLE>Association of Southern African Indexers and Bibliographers</TITLE>
</HEAD>
```

Remember that this title appears on the top of the browser, and not on the page itself; the title that will be displayed on the web page appears in the BODY section of the HTML document.

Each page in a website should have its own set of keywords and descriptions. It is however desirable to use the name of the website as the title for all the pages, along with subtitles for specific pages. This will make the web page more identifiable to both users and search engines (Rowland 2000b:3).

Different kinds of meta-tags are available to identify properties of a document and assign values to those properties. The title, keyword and description meta-tags are the most important ones that are used by search engines for indexing purposes. Title and author metadata can often be generated automatically from web pages. Subject keywords (words describing the subjects of web pages) and descriptions (abstracts or summaries of the content for display by search engines) usually have to be generated by people (Browne & Jermey 2001:8). Alimohammadi (2003:239) explains that meta-tags should always be placed at the head of the HTML document, just after the <TITLE> element, between the actual <HEAD> tags and before the <BODY> tags. The following example illustrates what the meta-tags for the ASAIB website could look like:

```
<HEAD>
<TITLE>ASAIB</TITLE>
<META NAME = "author" CONTENT ="Marlene Burger">
<META NAME ="description" CONTENT ="Official web site for Association of Southern
African Indexers and Bibliographers (ASAIB). It includes the personal details for
the executive committee members as well as the Association's training courses,
conferences and publications. It also includes a directory of freelance indexers who are
members of ASAIB">
<META NAME = "keywords" CONTENT = "book indexing, periodical indexing, indexing
workshops, freelance indexers directory, indexing conferences">
<HEAD>
```

Note that keywords can also be included in the <TITLE> tags. One could for example include a business slogan or a few keywords with the name of the company in the <TITLE> tag.

Description Meta-tag

The description meta-tag will be displayed when a user enters a query and the search engine retrieves some of the relevant websites. Alimohammadi (2003:240) advises indexers to use a description meta-tag if there is no descriptive paragraph at the start of the document, but to omit it if there is one. It is particularly useful for documents with a small amount of text, such as one consisting mainly of photographs.

Search engines usually take the first two lines of text, or the first number of characters (the exact number of characters is determined by the search engine) found on a website, and use them as the description. Those lines might not necessarily clearly describe what the website is about. The description included in the description meta-tag should be short, concise and to the point, but should not be so compressed that it cannot appropriately reflect the contents.

Keywords Meta-tag

The keywords meta-tag helps search engines to categorise websites and allow people to find web pages more quickly. Without this meta-tag, search engines will choose words for the site from the title and text of the site and these might not be the words or terms that best describe it. It is important to use only those keywords that the user might type in a search engine to try to find the website. One should also think of acronyms, synonyms, related words or word combinations that could be used. It is advisable to keep keywords in lower case since most users do not capitalise most proper nouns when searching on the Internet. Names are traditionally capitalised and should be treated this way when used as keywords (Du Preez 2002:119).

Be aware that most search engines do have limits on how many keywords are viewed, so make sure those that are used are as concise and as specific as possible (Alimohammadi 2003:241).

There is the problem of misuse with the keywords meta-tag. It is called 'spamindexing', 'spamdexing' or 'spamming'. Misuse includes repeating keywords over and over or using words that really have nothing to do with the website but are of interest to people, for example 'sex'. Keep in mind that although the misuse of keywords will increase traffic to the website, hate mail will also increase. Most larger search engines now also counter spamming strategies by ignoring the entire tag (Alimohammadi 2003:241).

Conclusion

The purpose of online indexing or web indexing was explained throughout this chapter. Attention was paid to indexing pitfalls when embarking on an online or web indexing project. Some web indexing questions and considerations were also answered. Indexing

both as a navigational tool and for information retrieval (IR) was then discussed. These included discussions on HTML coding and the use of meta-tags to assist search engines in accurately indexing the website.

BIBLIOGRAPHY

Alimohammadi, D. 2003. Meta-tag: a means to control the process of web indexing. *Online Information Review*, 27(4):238-242.

Broccoli Information Management. 2003. *Web site indexing.* Available: www.bim.net/website_indexing.html (accessed on 4 December 2003).

Broccoli, K & Van Ravenswaay, G. 1999. Web indexing, anchors away, in *Beyond book indexing*, edited by D Brenner and M Rowland. Phoenix: American Society of Indexers:37-42.

Browne, G & Jermey, J. 2001. *Website indexing: enhancing access to information within websites.* Adelaide: Auslib.

Cleveland, DB & Cleveland AD. 2001. *Introduction to indexing and abstracting.* 3rd ed. Englewood: Libraries Unlimited.

Denenberg, R. 1996. *Structuring and indexing the Internet.* Available: http://www.loc.gov/z3950/agency/papers/italy.html (accessed on 4 December 2003).

Desmontils, E & Jacquin, C. 2003. *Indexing a web site with a terminology oriented ontology.* Available: http://www.sciences.univ-nantes.fr/irin/indexGB.html (accessed on 4 December 2003).

Du Preez, M. 2002. Indexing on the Internet. *Mousaion*, 20(1):109-122.

Indexing the web. 2003. American Society of Indexers. Available: http://www.asindexing.org/site/webindex.shtml (accessed on: 4 December 2003).

Lathrop, L. 2003. *Consideration in indexing online documents.* Available: http:///www.bwa.org/articles/considerations_in_indexing_online_documents.htm (accessed on 4 December 2003).

Maislin, S. 2000. Ripping out the pages, in *Beyond book indexing*, edited by D Brenner and M Rowland. Phoenix: American Society of Indexers:43-57.

Maislin, SA. 2003. *What exactly is 'online indexing'?* Available: http://www.taxonomist.tripod.com/indexing/paperless.html (accessed on 4 December 2003).

Rowland, MJ. 2000a. <META> tags, in *Beyond book indexing*, edited by D Brenner and M Rowland. Phoenix: American Society of Indexers:71-76.

Rowland, MJ. 2000b. Website indexing: prepared for the Web Indexing Workshop, presented at the June 1999 Annual Conference of the American Society of Indexers and updated July 2000. Marisol Productions. Available : http://www.marisol.com (accessed on 4 December 2003).

Sullivan, D. 1998. *How to use meta tags.* Available: http://searchengingwatch.com/webmasters/article.php/2167931

Wright, JC. 1997. How to index online. *The Indexer*, 20(3):115-126. Also available: http://www.mindspring.com/~jancw (accessed on 4 December 2003).

12 METADATA AND INDEXING

Ansie Watkins

Abstract

This chapter introduces the reader to the concept of metadata and its purpose, and gives an overview of the various types of metadata. The importance of standards is emphasised, with a brief discussion of the most frequently used standards with regard to metadata and the inter-operability of systems. The Dublin Core as a widely used metadata standard is discussed in some detail. Finally, the steps of creating metadata during a digitisation project are discussed.

What is Metadata?

Librarians have been creating metadata since the earliest times, when they documented lists of items such as clay tablets, parchment scrolls and handwritten manuscripts in their collections. The term 'meta' originates from a Greek word meaning 'with, next, after, alongside' and, more recently, in Latin or English it refers to something beyond nature. Metadata can be described as data about other data, in this case information resources. It describes the characteristics of these information resources that can be in various formats but are mostly associated with networked information resources on the Internet. The activity of creating metadata may be compared with cataloguing and indexing that take place in libraries, archives and museums.

Why Metadata?

The purpose of metadata is to describe the content of an information resource (descriptive metadata). Apart from content description metadata is also used for administrative control, security, personal information, management information, content rating and resource identification.

To be effective for organisations that interact, metadata must be structured and consistent in its application. The ideal is therefore that a metadata standard is widely applicable. Metadata provides a link between the creator and the user of the information. It adds value to networked electronic resources. It enhances the retrieval process by enabling potential users to find relevant information easily and to exclude irrelevant information from their search results.

The term aims increasingly at the identification, location and description of these resources and to facilitate inter-operability of various electronic collections, such as digital images

of artworks, electronic texts of literary works, digital collections of research materials, electronic books and journals and many others. These resources are created by several different organisations, such as publishers, graphic designers, photographers, universities and digital libraries.

A metadata record consists of a set of predefined elements that represent specific attributes of an information resource and each element can have more than one value. This record is associated with an electronic resource that can be in various formats. The set of elements may exist as a separate record that is linked with the electronic resource or as part of the electronic resource itself.

How do the Concepts 'Metadata' and 'Indexing' Relate?

Indexing of content description is important in any digital project. Creating an index of values to describe digital resources is a vital component of the digitising process. Effort required to produce an index varies, depending on the nature of objects and depth of retrieval required. At some point in the process, human intervention is needed to assign values to the index terms. The index can be built after the conversion of an object into digital form.

Assigning index terms will add value to searching and retrieval of digital resources. All the traditional indexing practices and principles also apply in the case of indexing digital resources. These index terms will be included in the elements of the metadata structure that will be used to describe the electronic resources. The metadata structure is pre-defined and this process will be discussed later on in this chapter.

As in the case of 'traditional indexing' it is also necessary to decide on an indexing standard and format for the indexing of digital objects. Consistency is important to achieve valid data. Several indexing models have been developed for indexing digitised objects. The most commonly used indexing standards are MARC cataloguing and the Dublin Core metadata scheme.

Different Levels and Types of Metadata

One may distinguish different types of metadata, all of which perform different functions but which are often interrelated. Descriptive, structural and administrative metadata may be created at collection or item level.

Levels of Metadata

Collection-level Metadata

Collection level metadata refers to data about collections of information resources. It is used for a holistic description of a collection, such as archival collections or journals.

Item-level Metadata

This level of metadata describes individual items within a collection, for example photographs or correspondence in an archival collection, and articles in an electronic journal.

Different Types of Metadata

Descriptive (Content) Metadata

This type of metadata refers to the bibliographic and subject description of an entity. This type is the most familiar to us, for example MARC, Dublin Core, TEI (Text Encoding Initiative) and EAD (Encoded Archival Description).

Structural Metadata

Structural metadata is information that is relevant to the presentation of the digital object to the user, describing it in terms of navigation and use. Typical examples of structural metadata elements are

- unique identifier
- content type (text, digital image, file formats – XML, jpg, pdf)
- data file size
- structural divisions (which indicate various parts of the document, e.g. letters in a diary)

The University of California, Berkeley, defined an excellent example during the Making of America II project.

Administrative Metadata

Administrative metadata allows the repository to manage its collection, and includes the following:

- data related to the creation of the digital image (date of scan, resolution, etc.)
- data that can identify a version or edition of the image and help determine what is needed to view or use it
- ownership, rights and reproduction information

Some metadata elements may be both structural and administrative and may be used for similar purposes. Administrative metadata is critical for long-term management. Without well-designed administrative metadata, a file may be unrecognisable and unreadable within a decade. Examples of the elements in administrative metadata include

- source type – to identify the material from which the digital file was created
- source physical dimensions
- source ID – a local catalogue for a book or accession number for a special collections item
- scanning date
- scanning resolution
- owner of the copyright
- copyright date
- distribution restrictions

Metadata Standards

Literally hundreds of metadata standards have been created and are used throughout the world. The reason for this is to accommodate information resources that exist in several diverse

forms. To be effective, metadata must be structured. Another requirement for metadata to be effective is that it should be widely applied in a consistent manner. The goal of the creator of metadata should be to choose a popular and persistent standard that offers inter-operability with other metadata databases and one that is simple enough for public use.

A few examples of the most frequently used metadata schemes that became international standards for descriptive metadata have been mentioned above. The TEI and EAD will be briefly described later in the chapter. The Dublin Core will be discussed in more detail to illustrate how metadata is created and applied in practice.

Administrative, technical and structural metadata are more often uniquely defined by the institutions who created or are hosting collections of digital resources and aimed to meet their specific requirements. These elements are not necessarily displayed online because there may be no need in this regard. It is, however, important that once these metadata sets have been defined they should be consistently applied. The use of controlled vocabularies can enhance the consistent retrieval of relevant resources.

Something about Syntax

Data will not be usable unless the encoding scheme understands the semantics of the metadata scheme. Apart from the metadata sets there are therefore also a few other standards related to metadata that should be mentioned, namely those of syntax. This refers to the language or semantics in which the metadata set has been encoded. Encoding plays an important role in the creation of metadata. It differs according to the type of document. Examples of international standards in this regard are as follows.

Standard Generalised Markup Language (SGML)

SGML is a non-proprietary language for describing highly structured information. It is very useful for detailed encoding of information in databases and also for the conversion between different formats. It is rather complex, requires a high level of technical expertise to use it effectively and 'raw' SGML is not supported by web browsers.

Hypertext Markup Language (HTML)

HTML is the standard text formatting language for documents on the Internet and much simpler to use than SGML. It consists of text files encoded with tags that tell the computer how to format text on the screen, but is not recommended for highly structured information. HTML tags can also be used to encode metadata. *HTML Meta-Tags* is an initiative of meta-tagging web pages in order to make them searchable by means of search engines. Consult the following websites for more details about meta-tags:

How to use HTML
Meta_Tags http://searchenginewatch.com/webmasters/article.php/2167931
A Dictionary of HTML
Meta-Tags vancouver-webpages.com/META/**metatags**.detail.**html**

eXtensible Markup Language (XML)

eXtensible Markup Language (XML) is a subset of SGML that has emerged as a widely used international text-processing standard. It is as powerful and flexible as SGML, but not as complex and also supported by web browsers. It is selected by many international initiatives as an encoding standard and is used for several metadata schemes.

The following Internet links can be consulted for more detailed information on XML:

eXtensible Markup Language (XML)
http://www.w3.org/XML/
Sgml/Xml Based Metadata Initiatives
http://www.ifla.org/II/metadata.htm#sgmlxml

Resource Description Framework (RDF)

The RDF is a basic language, developed by the Worldwide Web Consortium (W3C) for the writing of metadata and processing of it on the Internet in a flexible way. Therefore various institutions can define and use metadata according to their needs, but also according to international standards. XML is used to write RDF tags.

More information on the RDF developments, including technical documentation, is available at *W3C Metadata area*: http://www.23.org/Metadata/ (Dublin Core Usersguide Glossary http://library.csun.edu/mwoodley/dublincoreglossary.html)

Examples of Metadata Standards

As mentioned before several metadata schemes are used throughout the world to describe different types of information resources. A few well-developed examples will be discussed in this section.

Text Descriptors: Text Encoding Initiative (TEI)

The Text Encoding Initiative is an ongoing international research project that has developed an extremely flexible set of guidelines, mostly for describing academic electronic texts, such as literature and reference works. It may be used for a single work or a collection of single works.

A TEI document consists of a bibliographic header and the text. Both these components are encoded in great detail. XML is increasingly being used as the encoding language, but many TEI documents are still in SGML. TEI documents can be created by means of XML editors and indexed by means of search and display software.

An explanation of how to use the TEI Guidelines to create electronic text will take up a reference source on its own. Therefore more detailed information on the TEI may be accessed on the following website:

The Text Encoding Initiative
http://www.tei-c.org/P4X/

Encoded Archival Description (EAD) for Archival Finding Aids

The EAD had its origin at the University of California, Berkeley. It was developed as a non-proprietary standard for machine-readable archival finding aids, namely inventories, registers, indexes and other documents created by archives, libraries, museums and repositories to support the use of their holdings:

- The information in a finding aid describes, controls and provides access to other information – not itself.
- It preserves and enhances the current functionality of existing inventories.
- The standard is intended to facilitate interchange and portability.
- The needs of public users, curatorial and reference staff were given priority in the standard's design.

The finding aid consists of the following segments:

- information about the finding aid itself
- a 'front matter' or title page
- the actual finding aid
- hierarchically organised information
- adjunct information that facilitates the use of the papers by researchers (e.g. a bibliography)

The EAD header is based on the TEI header and consists of

- EAD identity
- file description
- profile description
- revision description
- footer

Other Examples of Metadata Standards

These are as follows:

- music (Standard Music Description Language)
- images and objects descriptors (Categories for the Description of Works of Art, Consortium for the Computer Interchange of Museum Information [CIMI])
- classification (CyberDewey)
- numeric data (Standard for Survey Design and Statistical Methodology Metadata [SDSM])
- geospatial data (Content Standards for Digital Geospatial Metadata)
- MARC (Machine Readable Cataloguing)

Inter-operability of Metadata

The most important requirement for success of metadata standards is that it should be inter-operable and widely applied in a consistent manner. Metadata harvesting and metadata crosswalks will be discussed as two concepts that are associated with the inter-operability of metadata that will enhance accessibility and availability of networked electronic resources.

Crosswalks

A metadata crosswalk is a table that maps relationships between two or more metadata formats. It supports the ability of search engines to search effectively across different databases; in other words it promotes inter-operability.

The Library of Congress has prepared a useful article on MARC to Dublin Core Crosswalk that can be viewed at http://www.loc.gov/marc/marc2dc/html

Metadata Harvesting

The Open Archives Initiative (OAI) is an example of how metadata can be used to add value to electronic information resources. It develops and promotes inter-operability standards that aim to facilitate the efficient dissemination of content by means of a process that is known as metadata harvesting. A harvester collects metadata from various repositories. A repository is a server that is accessible for collection metadata. See http//www.openarchives.org/documents/FAQ.html

Dublin Core Metadata

What is the Dublin Core?

The Dublin Core metadata standard is a very effective element set for the description of a wide range of networked electronic resources. It is an example of descriptive metadata, but also contains elements that can serve as structural or administrative metadata. The Dublin Core also recommends a wide range of controlled vocabularies that can be applied to various elements.

See http://dublincore.org/documents/usageguide/

Advantages of the Dublin Core

The Dublin Core is often criticised for not making provision for all the types of information that people would like. But its purpose is to merely serve as a lowest common denominator for describing electronic resources. The advantages of the Dublin Core are as follows:

- It is usable and flexible.
- Its semantics are simple enough to be understood by a wide range of users without intensive training.
- The elements are easily identifiable.

- It is not intended to replace other resource descriptions, but rather to complement them. Other important metadata such as archival and accounting data were deliberately excluded to keep it as simple and usable as possible.
- It is mostly syntax-independent to support the widest range of applications.
- All elements are optional, but allow each site to define which elements are mandatory and which are optional.
- All elements are repeatable.
- The elements may be modified in limited and well-defined ways through the use of specific qualifiers, such as the name of a thesaurus used in subject qualifiers.
- It can be extended to meet demands of more specialised communities.
- Elements can be added for site-specific applications and it could be mapped to a more controlled system such as MARC.
- Dublin Core has received wide acceptance among the electronic information community and is seen as the unofficial Internet metadata standard.

See 'An Introduction to Metadata' at http://www.itb.hu/fejlesztesek/meta/hgls/core/Background/An_Introduction_to_Meadata.htm

The Dublin Core Element Set

The Dublin Core standard includes two levels: Simple and Qualified. Simple Dublin Core consists of 15 elements. Qualified Dublin Core has an additional element, namely Audience. It also contains qualifiers that describe the elements in more detail and may be useful during the retrieval process.

The Dublin Core elements can be defined in three categories:

- Elements related to the content of the resource:
  - title
  - description
  - subject
  - source
  - language
- Elements related to intellectual property:
  - creator (person/organisation responsible for intellectual or creative content)
  - publisher
  - contributor
  - rights
- Elements related to the physical manifestation:
  - date
  - type
  - format
  - identifier

Each element is optional and repeatable.

Dublin Core Metadata Template

Dublin Core Element Set Version 1.1- Reference Description http://dublincore.org/documents/dces/

The following table defines the various elements with brief comments on how they are used and also on standards and controlled vocabularies that may be used to qualify the elements. See the Dublin Core website for more information.

Elements	Definition	Comments
Title	A name given to the resource	Typically, a Title will be a name by which the resource is formally known
Creator	Person/Organisation responsible for creating the intellectual or creative content of the resource	Examples of a Creator include a person, an organisation or a service. Typically, the name of the Creator should be used to indicate the entity.
Subject	The topic of the content of the resource	Typically, a Subject will be expressed as keywords, key phrases or Classification Codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.
Description	An account of content of the resource	Description may include (but is not limited to) an abstract, table of contents, reference to a graphical representation of content, or a free-text account of the content.
Publisher	An entity responsible for making the resource available to public	Examples of a Publisher include a person, an organisation, or a service. Typically, the name of a Publisher should be used to indicate the entity.
Contributor	An entity responsible for making contributions to the content of the resource	Examples of a Contributor include a person, an organisation, or a service. Typically, the name of the Contributor should be used to indicate the entity.
Date	A date associated with an event in the life cycle of the resource	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [w3CDTF] and follows the YYYY-MM-DD format.
Type	The nature or genre of the content of the resource	Type includes terms describing general categories, functions, genres or aggregation levels for content. Recommended best practice is to select a value from controlled vocabulary (e.g. the working draft list of Dublin Core Types [DCT1]) To describe the physical or digital manifestation of the resource, use the format element.

Elements	Definition	Comments
Format	The physical or digital manifestation of the resource	Typically, Format may include the media-type or dimensions of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (e.g. the list of Internet media types [MIME] defining computer media formats).
Identifier	An unambiguous reference to the resource within given context	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. For example formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator [URL]), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).
Source	A reference to a resource from which the present resource is derived	The presence of the resource may be derived from the Source resource in whole or in part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.
Language	A language of the intellectual content of the resource	Recommended best practice for the values of the Language element is defined by RFC 1766 [RFC1766] which includes a two-letter Language code (taken from the ISO 639 standard [ISO639]), followed optionally, by a two-letter Country code (taken from the ISO 3166 standard [ISO3166]). For example, 'en' for English, 'fr' for French, or 'en-uk' for English used in the United Kingdom.
Relation	A reference to a related resource	Recommended best practice is to reference the resource by means of string or number conforming to a formal identification system.
Coverage	The extent or scope of the resource	Coverage will typically include spatial location (a place name or geographical coordinates), temporal period (a period label, date or date range) or jurisdiction (e.g. named administrative entity).
Rights	Information about rights held in and over the resource	Typically, a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright and various Property Rights. If the Rights elements is absent, no assumptions can be made about the status of these and other rights with respect to the resource.

The Nordic Metadata Project offers an online template for creating Dublin Core metadata. If you use the metadata created by this form and follow the examples, term lists and recommendations, your HTML documents will carry high quality metadata. Please consult the website for more details: http//www.lub.lu.se/cgi-bin/nmdc.pl

Creating Metadata

Process of Creating Metadata during a Digital Project

Metadata may be deployed in various ways, such as

- embedding metadata in a web page – metatags in HTML encoding
- separate HTML document linked to the resource it describes
- in a database linked to the resource

Descriptive metadata for digitised documents are often found in an external database. In other words the index terms and their values are stored in a separate database from the images. Once images have been created and stored in a database, discovery and retrieval become critical functions. Indexing each image is necessary to ensure the location of images during a search. The digital resource can only be useful if it contains metadata to allow users to navigate the digital file.

The steps in creating metadata during a digitising project can be summarised as follows:

- Select a suitable descriptive metadata scheme.
- Establish administrative metadata.
- Decide on structural metadata.
- Decide on technical standards for digital images, electronic texts, file formats, software tools, and so on, to create digital resources and metadata templates.
- Establish the naming convention for a set of filenames for unique identifiers and match with corresponding record.
- Set filenames, attribute fields (elements), order, checklist.
- Develop indexing protocols to ensure consistency.
- Decide on controlled vocabularies if applicable.
- Develop a quality control system.
- Develop metadata template to create metadata records.
- Train staff to create digital resources (if applicable), apply indexing principles and use the templates and other software tools as needed.
- Create digital resources.
- Create metadata records by assigning values to the pre-defined elements.
- Run the automated indexing program to enable the search engine to retrieve the records during an online search.
- Create or customise a user-friendly interface that will enable clients to search and view electronic resources online.

- Evaluate the system.
- Market the system.
- Teach clients to use the system.

Tools for Working with Metadata

Several software tools are available for the following:

- creating metadata
- creating/change of templates
- automatic extraction/gathering of metadata
- converting between formats
- integrated (tool) environments
- formatting and reviewing tools
- XML tools
- publishing tools
- searching tools

Tools can be downloaded from

Metadata UKOLN Software tools: http://www.ukoln.ac.uk/metadata/softwaretools/

Dublin Core Metadata Initiative (DCMI) Tools and software: http: dublincore.org.tools

Tools for working with metadata: http:///www.csc.noaa.gov/metadata/metatools.html

The IFLA website offers a comprehensive and detailed list of metadata resources: http://www.ifla.org/II/metadata.htm

Cataloguers Toolbox: http://staff.library.mun.ca/staff/toolbox/standards.htm

Conclusion

The Internet is an extremely dynamic environment. Therefore the optimisation of access to Internet resources offers a challenge to librarians and information workers. They should become proactively involved in the description of resources from which their clients will benefit. The ongoing increase in the creation and use of networked electronic resources will impact on the demand for metadata. Metadata will become increasingly important as we develop library systems that are inter-operable.

Although the few standards that were discussed here are well tested and widely used, this should not prevent librarians from exploring new opportunities in the field of metadata. Information resources are continually created in new or improved formats. Metadata standards should keep up with new developments. The guidelines in this chapter are based on resources prepared by international initiatives in the field of metadata. References as indicated throughout the text are constantly updated and should be consulted on an ongoing basis.

13 ABSTRACTING

Marlene Burger

Abstract

The abstract (produced by machine or human), as a concise statement of the central message of a document, has become an increasingly important tool for distinguishing truly relevant information from the bulk of information available (Pinto & Cálvez 1999:365).

Abstracting and Abstracts

Abstracting is a verbal method of organising information (i.e. words are used to describe what an entity is about). It is aimed at making the mass of information more manageable so that clients can retrieve relevant information when they need it, and exclude irrelevant information. This means that, when you are writing abstracts for your own information system, you should be guided by your own needs. To be able to do this properly, you should be familiar with the general rules and guidelines that apply to the writing of abstracts.

An abstract is a short, concise and accurate representation (surrogate) of the intellectual content of an entity, its purpose being to promote awareness of the entity and also of information retrieval. Abstracts can be written for entities such as books, reports, journal articles and websites. In this chapter we concentrate on abstracts for journal articles, but the principles can be applied to other types of entity as well.

There are five types of abstract: indicative abstracts, informative abstracts, critical abstracts, structured abstracts and modular abstracts. Diagrammatic abstracts are not discussed. The types of abstract most commonly used are indicative and informative abstracts. Frequently, a combination of indicative and informative abstracts is used.

Indicative Abstracts

Short abstracts which describe the broad concepts covered in an entity in general terms are called indicative abstracts. General terms like 'are discussed' and 'was investigated' are used to describe the intellectual content of an entity. The purpose of indicative abstracts is merely to show that the information exists: no factual information or findings are included. Indicative abstracts only indicate what is covered in the entity – what is done, what is discussed, what is analysed, what is compared, and so on. Indicative abstracts are presented in a single paragraph only.

Indicative abstracts indicate that important information and/or specific data may be found in a particular entity (e.g. the number of sheep farmers in South Africa is determined according

to types of sheep). Thus indicative abstracts focus attention on the existence of entities. They are considerably shorter than informative abstracts and vary from 50 to 100 words.

Example of an indicative abstract (adapted from Cleveland & Cleveland 1990):

> Disraeli, H. 1989. Onion raising in California. *Agriculture Happenings*, 201(2):7-19.
>
> Deals with all aspects of the onion crop in California and touches onion production in other states, mainly for comparative purposes. All types of onions are discussed, with figures given for each category and for the total production. There are breakdowns for how much of the product is consumed in the state and how much is exported. There is also a detailed discussion of farming techniques, wholesale practices, transportation, and general economics of the onion business. A large portion of the article is devoted to the social and medicinal implications of eating onions.

Informative Abstracts

Longer abstracts that summarise the essential factual content of an entity as concisely and clearly as possible are called informative abstracts. They include *qualitative* and *quantitative* information, as well as the following: the purpose and scope of the entity, the research methods used, the research results, the validity of findings and their applications. This depends on the extent to which the information is covered in the entity – for example not all entities are scientific articles in which these matters are set out step by step. Informative abstracts may be presented in more than one paragraph and vary from 200 to 300 words and may sometimes even be longer. The length of informative abstracts may also be restricted to a tenth or a twentieth of the original entity.

If the information is available, informative abstracts should include the following facts/points about the original entity:

- aim and scope of the entity
- methods of use
- results
- findings
- validity of findings
- applications

Informative abstracts take considerably longer to write than do indicative abstracts. The intellectual content of entities – what they are about – has to be thoroughly analysed first. Knowledge of the subject content is an additional recommendation.

In practice, there is a tendency to write a combination of indicative and informative abstracts (indicative-informative abstracts). This means that the chief aspects are discussed in informative style and the less important aspects in indicative style.

Example of an informative abstract (adapted from Cleveland & Cleveland 1990):

Disraeli, H. 1989. Onion raising in California. *Agriculture Happenings*, 201(2):7-19.

According to this article, California raised 170,000,019 onions while its closest competitors, Oregon and Texas, combined could manage only 70,000,017 onions. Rhode Island was almost out of contention with a little over 2 million. Big, white, sweet onions are the best sellers, with yellow onions second, and small green onions close behind. Some 120 million white onions, 30 million yellow onions, and 20 million little green onions were put on the market. Californians consumed 70,000,035 of these onions, with the rest being placed on the out-of-state market. Most onions farming is highly mechanical and is done by large-scale farming enterprises. The main mode of transportation is by truck, although some is by rail and a minute part by Ford station wagon. The onion business is economically stable and is generally a money-maker. According to the article, persons who eat onions with every meal have a life expectancy of some 1% above the national average, but they are invited to only 20% as many parties, and their divorce expectancy is 200% above the national average.

Critical Abstracts

Normally abstracts include no criticisms or personal interpretations. Critical abstracts are the exception. Critical abstracts are seldom found in databases with an international target group (e.g. *ERIC* or *Philosopher's index*). Critical abstracts will become increasingly important because of the role they can play in value addition by information services.

Critical abstracts comment, for example, on the

- depth or average of an entity (e.g. a journal article or research report)
- adequacy of the research methodology used and the research findings where applicable
- background needed by the target group (clients) to utilise the entity
- importance of the contributions to the development of knowledge

Although critical abstracts can add value to information services, they are very time-consuming to write and require extensive knowledge of the subject. The abstractor must include a *substantial opinion* about the article. Critical abstracts are similar to critical book reviews – they evaluate the quality of the work and indicate the weaknesses and strengths of the article. This can even include comparisons with other articles.

Example of a critical abstract (adapted from Lancaster 1998):

Pao, YC & Shy, DS. 1982. On the relationship between bulk modulus and relative volume of the lung during inflation-deflation maneuvers. *Journal of Biomechanical Engineering*, 104(2):136-142.

The article presents an equation relating the bulk modulus of the lung to the relative volume during inflation and deflation. The average bulk modulus of the lung was obtained by injecting air via a 6-mm-i.d. cannula in the main lobar bronchus. "Regional lobe" volume changes were measured by roentgen-videographically determined placement of 25 metal markers implanted in the excised lower lobes of three dogs. Whole lobe volumes at various transpulmonary pressures were measured by water

displacement. Pressure and volume measurements were used to calculate bulk modulus ($K = \Delta V P / \Delta V$). The "most satisfactory least squares curve-fit" of bulk modulus (K) vs. relative volume (V/V max) was obtained with the equation $K = C/(1 - V/V\ max)n$. Substituting for bulk modulus, with the equation $K = VdP/dV$, and integrating enabled computer-generated pressure-volume plots. This equation provided a better pressure-volume curve-fit than previously obtained, especially at low values pressure and volume. Also, as expected, the bulk modulus was smaller at low volume, but the rate of change of modulus was greater during deflation than inflation.

The authors assumed, without giving sufficient justification, that the "regional lobe" (the area bounded by the 25 markers) included a higher density of airways than the rest of the lobe. Using this assumption, the authors claimed that the modulus and the rate of change of modulus were different for parenchyma tissue and the airways during both inflation and deflation. No mention, however, was made of paired t-tests or any other statistical tests. In fact, if they had done a paired t-test, they would have discovered that none of these differences were significant, even at the 90 percent confidence level.
Other sources of error which were not addressed include: the difference in the properties of excised lung and intact lung due to blood in the vessels, surrounding tissue, negative pressure, etc; the effect of the markers on the pressure-volume relationship; the effect of strain rate on the modulus of lung tissue, which is a viscoelastic material; the time lapsed between regional volume measurement and whole volume measurements (this is important for viscoelastic material); the difference between the true regional ΔV and the measured ΔV; and the differences between the mechanical properties of dog and human lung tissue.

Despite its limitations, the article presents a step forward in the understanding of mechanical properties of the lung, and, thus, lung diseases. Therefore, it should be of benefit to researchers interested in respiratory mechanics and physiology.

DS Feldman, USA

Structured Abstracts

Structured abstracts have the same requirements as informative abstracts, but the content is written in separate paragraphs and under different headings. During analysis of the article, the abstractor must take note of any pre-determined aspects, which will form the headings for the presentation. The following is an example (Lancaster 1998) of a framework for a structured abstract for a specific article:

Irrigation type	Soil type	Crops	Climatic conditions	Place	Results

The abstract need not be presented in columns as illustrated above. The following example (adapted from Lancaster 1998) illustrates an alternative format:

BACKGROUND: Structured abstracts, containing several subheadings, have replaced traditional abstracts in most medical journals. Evaluation studies have shown that such abstracts usually provide more information, are of a higher quality, facilitate peer review and are generally welcomed.

AIM: The aim of the studies reported here was to investigate an additional possible advantage for structured abstracts, namely whether or not they are easier to search.

METHOD: Two studies are reported. In Study 1, using an electronic database, 52 readers were asked to find the answers to two questions for each of eight abstracts in one format (e.g. traditional), followed by two questions for each of eight abstracts in the other format. Time and error date were recorded automatically. In Study 2, using a printed database, 56 readers were asked to find five abstracts that reported a particular kind of study (e.g. studies with school children and reading tests) and then to find five more that reported another kind of study. Again, the order and presentation of the format of the abstracts was counterbalanced. Time and error data were recorded manually.

RESULTS: In Study 1, the participants performed significantly faster and made significantly fewer errors using the structured abstracts. There were, however, some unexplainable practice effects. In Study 2, the participants again performed significantly faster and made significantly fewer errors with the structured abstracts. In Study 2, however, there were asymmetrical transfer effects: participants who responded first to the structured abstracts responded more quickly to the following traditional ones than did those participants who responded first to the traditional abstracts.

CONCLUSIONS: The overall findings, notwithstanding certain caveats, support the hypothesis that is easier for readers to search structured abstracts than it is to search traditional ones.

Modular Abstracts

Modular abstracts are intended as full content descriptions of current entities. Each abstract consists of five parts: a citation, an annotation, an indicative abstract, an informative abstract and a critical abstract. This implies that an abstracting service can process it to conform to its own unique requirements with a minimum of effort: any one of the abstracts can be used intact, or edited to form, for example a partially indicative, partially informative abstract, or a partially informative and partially critical abstract. Modular abstracts are mostly accompanied by a modular index and together they represent the concept of 'modular content analysis'.

The prime purpose of modular abstracts is to eliminate the duplication and waste of intellectual effort involved in the independent abstracting of the same entities by several

services. Their standardised format (i.e. the five parts) and treatment also reduce repetitive handling and speed the flow of work within recipient abstracting services.

Example of a modular abstract showing the five parts (adapted from Lancaster 1998):

Citation
Rosensweig, RE & Beecher, N. 1963. Theory for the ablation of fiberglass-reinforced phenolic resin. *American Institute of Aeronautics and Astronautics Journal*, 1(8):1802-1809.

Annotation
A theoretical model is developed, for a charring and melting composite material, combining glassy ablation and the char layer-molten glass chemical reaction effects.

Indicative
The variables associated with the ablation of a typical resin-glass system are examined. These include glass ablation and plastic pyrolysis, flow in both the reacting and non-reacting parts of the melt, mass loss and heat absorption due to chemical reaction. The mathematical development is traced and the approximations utilised are discussed. Parametric examinations are made.

Informative
Pyrolysis, melting, and chemical reaction are taken into account in this theory of the ablation of phenolic-fiberglass. It postulates a very thin, isothermal, surface reaction zone, where the char layer (carbon) formed during the pyrolysis of the organic binder reacts chemically with the molten silica. Other assumptions are conventional. Calculations for typical IRBM re-entry conditions showed little temperature drop in the reaction zone, 6% maximum and usually less than 1%. Depth of the zone was three orders of magnitude less than the thermal thickness. The unreacting run-off in the melt zone ranged from 40-80% as a function of the possible reaction enthalpy level. However, more than 99% of the material reaching the reaction zone was affected. At the expected temperatures of 1400-2000 °C, the theory assumed the reaction SiO2 + 3C > SiC + 2CO. Earlier experiments had yielded the reaction kinetics. Significant effects, up to 25% increase, on the ablation rate appeared only at the lower reaction rates. Changing the reaction enthalpy by a factor of three changed the ablation rate by less than 10%. When compared with a peak re-entry ablation rate, the value given by this theory was reported to be 38% in defect.

Critical
This theory extends the classic work of Bethe and Adams (Avco-Everett Research Lab, Res Rept 38, Nov 1958) on ablation of pure glass. Thus it treats the problem as concerning carbon-contaminated glass rather than, as is more usual, a char-layer. In the only comparison given between the theory and experimental data, revealing 38% underprediction by the theory, a thorough error analysis was not included. Spalding (Aero Quart, Aug 1961, pp 237-274) and Scala (General Electric Co (MSVD), Rept R59SD401, July 1959, ARD Jnl, June 1962, pp 917-924) have tested similar problems.

The accompanying modular index entries suggest descriptive terms, drawn from representative index vocabularies that could be used intact, with refinement or with augmentation to index the abstract derived from the modular package. The representative index vocabularies

used as sources for the modular index entries are to be derived from the current indexes or authority lists of the recipient abstracting and indexing services, and will thus reflect the indexing styles and policies of these services.

Example of the modular index entries (adapted from Lancaster 1998) for the modular abstract illustrated above:

Physical and Mathematical Systems
Axisymmetric and Blunt Body Systems
Re-entry Bodies

Environment
Atmospheric Entry
Re-entry Conditions
Space Flight

Mass Transfer
Ablation, Analytical
Ablation, Charring
Ablation, Melting
Ablation of Glasses
Chemical Reaction Effects
Gasification Ration
Reaction Zone
Thermal Thickness

Authors
Beecher, N
Rosensweig, RE

Thermodynamics
Carbon-Silica Reactions
Coupled reactions

Materials
Ablation materials
Glass Fibers
Phenolic Resin
Phenolics, Fiberglass Reinforced
Reinforced Plastics
Rocket and Missile Materials
Thermal (Re-entry) Shields

Means and Methods
Parametric Analysis

Affiliations
Massachusetts Institute of Technology
National Research Corporation

Reasons for Writing Abstracts

Authors write abstracts for their own articles, reports, dissertations and theses. You will also encounter abstracts in many of the bibliographic databases, such as *ERIC* and *Library literature*. Abstracts are also used for some full-text databases to enhance the specificity of information retrieval. The reasons why you would write abstracts for your personal information system will be somewhat different from the reasons why abstracts are written as part of publications or for databases intended for general use. For large, general databases, abstracts are written for the following reasons:

- Abstracts make it possible to select entities (e.g. journal articles). You can use the contents of the abstract to decide whether it is worth tracing the complete entity.
- They make clients aware of the existence of specific information (e.g. new research findings, new products or new techniques).
- Abstracts can stimulate new ideas and new interest.
- Well-written and complete informative abstracts can save a great deal of reading time by giving a concise version of the essential facts covered by an entity.
- Abstracts could overcome language problems in instances where entities are available in foreign languages but at least the abstracts are in English. Sometimes an informative abstract may even serve as a substitute if the original entity is not available.
- Abstracts help with the retrieval of information because they offer more search points.
- Abstracts that form part of an entity can help indexers to decide on indexing terms, since these already give a good indication of the contents of an entity.
- Abstracts make it easier to work through the mass of published information and to keep up with new developments.

Steps in the Writing of Abstracts

Although there are no absolute guidelines or instant recipes for how to set about writing an abstract, there are certain steps you ought to bear in mind when writing indicative and informative abstracts.

1. First study the entity to determine what it is about. This is called *conceptual analysis* and is sometimes a difficult process. There is also no unanimity as to precisely what you should take note of. The complexity of conceptual analysis is covered in detail in chapter 3.
    - Scan the entity at least once to get an idea of what it is about.
    - Note the title, headings and subheadings.
    - Note the introductory and closing paragraphs.
    - Make notes on the key points.

2. On the basis of the conceptual analysis, decide which *key points* covered by the entity should figure in the abstract. Using this as a basis, you then write the abstract according to recognised standards and guidelines. This is referred to as the *translation* of the contents.

- Cover the aim and reason for the research, the scope, the research methods, and the results and findings, if available.
- Compare the points you identified in the conceptual analysis with the reason why the abstract is being written.
- Exclude negative information, generally familiar historical information, points that have not been researched, references to the works of other authors, tables and figures.
- Do not include criticism or your own interpretation.

3. Write the abstract, taking the guidelines into account.
   - Write the abstract in draft form, referring to the notes you made in step 2.
   - The first sentence should convey the nucleus of the intellectual scope of the entity without repeating the title.
   - Check whether you have to write an indicative or informative abstract, and whether it would be better to write a combination.
   - Note the recommended style for abstracts.
   - Note the recommended length.
   - Edit the abstract and note the punctuation marks, spelling errors, accuracy, omissions and conciseness. Accuracy is vital.
   - Write the final abstract.
   - Write an accurate bibliographic description if the abstract does not appear with the original entity. If possible, ask a more experienced abstractor to check your abstract.
   - Sign the abstract with your initials.

Guidelines for Writing Abstracts

There are standards that provide guidelines for the writing of abstracts. For example the following publication:

> National Information Standards Organisation. 1994. *Writing abstracts: a draft American national standard*. Bethesda, Md: NISO.

Individual compilers of databases such as *ERIC* create their own guidelines for abstracting according to such standards.

Bibliographic Description

If the abstract does not appear with the original entity, as in a journal article, it must have an accurate bibliographic description. If the bibliographic description is not accurate, it is very difficult and sometimes even impossible to trace the original entity. There are a variety of guidelines in the subject literature on the format that could be used. Different databases and/or journal publishers may prescribe different formats; but for the purpose of this chapter we follow the Harvard method, since it is an international standard. Keep in mind that most guidelines work with the same data elements even though their order of presentation may differ. These data elements are as follows: title, author(s), publication details, date, page numbers. Examples according to the Harvard method:

DDC21: keeping pace with knowledge. 1996. *Library Association Record*, 96(2):291. Also available online: OCLC Forest Press DDC Web Site http://www.oclc.or...22ddc/9722book.htm

Tenopir, C & Jascó, P. 1993. Quality of abstracts. *Online*, 17(3):44-45.

Zin, C. 2000. Success, a structured search strategy: rationale, principles, and implications. *Journal of the American Society for Information Science*, 519(13):1232-1247.

Please note that in some databases the titles of periodicals have standardised abbreviations, that is the full titles are not used. The bibliographic description usually precedes the abstract, appearing at the head of the abstract.

Abstracting Style

Abstracts should be written in compliance with the style prescribed for the type of abstract, for example indicative or informative abstracts. This implies that one should be able to recognise the type of abstract from the style used.

Writing Style

Although there are no fixed writing style guidelines, you may heed the following:

- The first sentence of the abstract should give a clear explanation of the intellectual content of an entity, without repeating the title.
- As far as possible, an indicative abstract should always be written in a single coherent paragraph (more paragraphs if informative or critical abstracts).
- It should be written as briefly and concisely as possible – a maximum amount of information should be conveyed with the minimum of words.
- Although you should avoid long, complex sentences, telegram style is not acceptable either. All sentences must be complete. Verbs, articles and prepositions may not be omitted. The first sentence may begin with a verb, for example 'Describe the state of indexing in South Africa'.
- Use only well-known abbreviations and acronyms. When other abbreviations or acronyms are used for the first time in the abstract, the words must be written out in full followed by the acronym or abbreviation in brackets, for example computer-assisted instruction (CAI). Then, in the rest of the abstract you may use the abbreviation or acronym.
- Any unnecessary words that do not convey significant information should be avoided, such as 'this article' and 'the authors made a study of'.
- As far as possible, the writing style and the sequence of the presentation should correspond to that of the author(s). Be aware that opinions about this do vary and that some authors feel that an abstract should be an independent intellectual creation (this may be achieved when writing a critical or structured abstract).
- Expressions such as 'I found' or 'we found' should be avoided. Rather use phrases such as 'findings show' 'research indicates' or 'the findings are'.

- There are no definite guidelines on the use of the past or present tense. You should be guided by what sounds best.

The following is a list of possible words and phrases that can be used when writing abstracts:

Acknowledges
Addressed by
Addresses the problem/issue of
Advances arguments for/against
Advises
Advocates
Analyses
Argues that
Asserts that
Assesses

Brief discussion of
Briefly mentions/looks at

Challenges
Classifies
Commands
Comments on
Compares the … with …
Concentrates on
Concludes that
Critically evaluates
Criticises

Debates the
Defines the concepts
Delineates
Describes
Differentiates between
Discusses
Distinguishes

Elucidates
Emphasises
Examines
Explains
Explanation of
Explores
Expresses concern for

Focuses on
Forecasts future developments in
Foresees

Highlights

Identifies
Illustrates
In-depth discussion of
Includes
Introduces
Investigates

Launches a strong attack on
Lists the winners of
Literary criticism of

Maintains that

Obituary of
Offers guidelines for
Outlines
Overview of

Pays tribute to
Pleads for
Popular discussion
Profile of

Questions whether/the
Quotes

Reasons for
Recommends that
Reflects on
Refutes
Regrets
Rejects
Remarks on
Reprint of extracts from
Review
- Review of a book which
- The reviewer contends/implies/maintains that
- The reviewer describes the book as

- Applicable only to review articles

Scrutinises
Shows how
Studies
Study of
Suggests
Summarises
Summary of
Surveys

Tabulates data on
Traces the career/history of

Wide-ranging discussion on

Bellardo (1991:34) gives the following examples to promote the economical use of words:

Phrase	Suggested alternative
because of the fact that	because
consensus of opinion	consensus
for the purpose of	for
in the event that	if
at this point in time	now
at that point in time/during the period of time that	then/while
in order to	to
in a similar manner	similarly
in view of the fact that	since/because

Accuracy

Accuracy is a very important requirement, especially in informative abstracts that may sometimes be used to replace the original entity. The accuracy of numerical data in particular and of names should be verified. For instance, there is a big difference between 1996 and 1696. If the author states that something ought to be researched or is going to be researched, you cannot say that it has already been researched. Nor should the author's emphasis of certain points be changed.

Comprehensiveness

An abstract must encompass all the main points covered in the entity. If the abstract is to be used for information retrieval, the aim and coverage of the information system, such as the database, should be taken into account. No concept or point contained in the entity should, however, be discussed in greater detail than was done by the author in the original entity. This also applies to the analysis of tables and graphs.

Omission of Information

The following information should preferably *not* be included in an abstract:

- References to other entities or contributions by other authors. Such references are permissible only if the original entity consists mainly of a discussion of another entity, or when you are writing a critical abstract.
- References to specific tables, graphs and figures. Remember that the table, graph or figure is not included in the abstract. This rule does not apply for critical or structured abstracts.
- Negative information such as 'the needs of information practitioners were not researched': research that has not been done is not mentioned.
- Criticism or interpretation. This may only be included in critical abstracts or other types of abstract with a special function.
- Historical or background information.
- Recommendations about future research.
- References to quotations in the original entity.
- Unfounded comments, well-known facts, speculation or information that does not appear in the original entity.

Signing the Abstract

Normally abstracts are initialled by the abstractor. In critical abstracts the abstractor's full name and credentials may be required.

Automatic Abstracting

Since the purpose of this chapter was to write about abstracts in general only, this section is but a brief discussion of automatic abstracting. In automatic abstracting computers are used to prepare extracts or summaries of the electronic text of entities. Lancaster (1998:267) states the following about automatic abstracting: 'If computers can be programmed to select terms from documents according to frequency criteria [as in automatic indexing], they can also be programmed to select sentences from documents.' This is the basis of what is usually called 'automatic abstracting' although it is more accurately referred to as 'automatic extraction'.

Automatic abstracting involves deriving meaningful sentences from the text of an entity, and compiling these sentences to reflect the intellectual content of the entity. Apart from word frequencies, the computer program also notes the relationships between the keywords in a sentence. (If you are interested in finding out more, please read the article by Luhn.) The style of an entity may also be used as a starting point in compiling abstracts. Here the first and last sentences of each paragraph are considered, or the first sentence of each paragraph only. Subject-oriented thesauri have also been used to identify important words in sentences. However, no real success has yet been achieved with automatic abstracting – it cannot replace human abstracting. Sometimes it supplies concise, understandable information, but too often the abstracts read like strings of disconnected sentences, which indeed they are. The main problem with abstracting by sentence extraction is that there is a lack of balance and cohesion. In experiments the following principles have been used:

- Sentences have a greater weighting if certain keywords appear.
- Certain words serve as cue words. If these words appear in a sentence, it implies that the sentence deals with a certain aspect.
- Words appearing in the title and headings of sections or subheadings, give a good indication of what an entity is about.
- Weighted words are assigned in accordance with the position of sentences. Sentences found in the introductory or concluding paragraphs will have a greater weighting than those appearing elsewhere in the entity.
- Sentences are rejected if certain words appear in them.
- Identifying sentences that are good indicators of the content, for example sentences that contain 'the main idea is' or 'a method is described'.
- Parsing of texts – the grammatical parts of a sentence are examined and identified.

The more formal the text, the easier it is to apply automatic abstracting. Johnson (1995:32) explains automatic abstracting as follows: 'The lack of coherency in automatic abstracts, where the reader is left to infer the relationship between the ideas expressed, detracts from the meaning conveyed and requires considerable time and effort for its understanding'. Much more research is therefore necessary to refine the process of automatic abstracting. A comparison between manual and automatic abstracting reflects the following (adapted from Lancaster & Warner 1993:274):

Manual abstracting: Select key sentences; paraphrase key concepts/sentences; organise into coherent paragraph; result: abstract.

Automatic abstracting: Select sentences by highest numeric score; modify sentences by linguistic analysis; organise sentences by rank statistical order or by sequence in which they occur; organise sentences by structural analysis; result: extract.

While Luhn (1958) and other authors used word phrase frequency in the selection of sentences, other investigators have proposed or used alternative criteria. Edmundson

(1969) identified four possible methods: the key method, the cue method, the title method and the location method. (For more information on these methods, please read the article by Edmundson.)

An example of a Luhn auto-abstract as it appears in Lancaster (1998:268) illustrates the key method:

Source: The Scientific American, Vol. 196, No. 2, 86-94, February, 1957.
Title: Messengers of the nervous system
Author: Amodeo S. Marrazi

Editor's Sub-heading: The internal communication of the body is mediated by chemicals as well as by nerve impulses. Study of their interaction has developed important leads to the understanding and therapy of mental illness.

Auto-abstract*

It seems reasonable to credit the single-celled organisms with a system of chemical communication by diffusion of stimulating substances through the cell, and these correspond to the chemical messengers (e.g., hormones) that *carry* stimuli from cell to cell in the more complex organisms (7.0)†

Finally, in the vertebrate animals there are special glands (e.g., the adrenals) for producing chemical messengers, and the nervous and chemical communication systems are intertwined: for instance, release of adrenalin by the adrenal gland is subject to control both by nerve impulses and by chemicals brought to the gland by the blood (6.4)

The experiments clearly demonstrated the acetylcholine (and related substances) and adrenalin (its relatives) exert opposing actions which maintain a balanced regulation of the transmission of nerve impulses (6.3)

It is reasonable to suppose that the tranquilizing drugs counteract the inhibitory effect of excessive adrenalin or serotonin or some related inhibitor in the human nervous system (7.3)

* Sentences were selected by statistical analysis when they had a degree of significance of 6 and over.
† Significance factor is given at the end of each sentence.

Kuhlen (1984:98) states that we do not know enough about the intellectual processes involved in abstracting to develop programs whereby this activity can be simulated by computer: 'Abstracting ... is an intellectual art and as such not directly transferable to automatic procedures. Cognitive psychology and artificial intelligence have, so far, not provided us with sufficient knowledge about the processes really going on in abstracters' minds when they understand texts and condense them. Thus the direct imitation of an intellectual procedure such as abstracting seems to be out of reach.'

A substantial amount of research and development in this regard has been undertaken during the past few years, but Kuhlen's words still seem relevant today.

Conclusion

While some progress has been made in applying computers to various tasks related to information retrieval, there is little evidence that automatic procedures can as yet outperform humans in such intellectual tasks as indexing, abstracting, thesaurus construction and the creation of search strategies. However, computers can be used to improve the productivity of abstractors and the quality of their work. It is most definitely an area where new developments should be monitored.

BIBLIOGRAPHY

Bellardo, T. 1991. *Subject indexing: an introductory guide.* Washington, DC: Special Libraries Association.

Cleveland, DB & Cleveland, AD. 1990. *Introduction to indexing and abstracting.* Englewood, Colo: Libraries Unlimited.

Craven, TC. 1993. A computer-aided abstracting tool kit. *Canadian Journal of Information and Library Science,* 18(2):19-31.

Cremmins, ET. 1996. *The art of abstracting.* 2nd ed. Arlington, Va: Information Resources.

Edmundson, HP. 1969. New methods in automatic extracting. *Journal of the Association for Computing Machinery,* 16(2):264-289.

Hartley, J. 1999. From structured abstracts to structured articles: a modest proposal. *Journal of Technical Writing and Communication,* 29(3):255-270.

How to implement a naturalistic model of abstracting: four core working steps of an expert abstractor. 1995. By B Endres-Niggermeyer ... et al. *Information Processing and Management,* 31(5):631-647.

Johnson, FC. 1995. Automatic abstracting research. *Library Review,* 44(8):28-36.

Kuhlen, R. 1984. Some similarities and differences between intellectual and machine text understanding for the purpose of abstracting, in *Representation and exchange of knowledge as a basis of information processes*, edited by HJ Dietschmann. Amsterdam: North-Holland:87-109.

Lancaster, FW. 1998. *Indexing and abstracting in theory and practice.* 2nd ed. London: Library Association.

Lancaster, FW & Warner, AJ. 1993. *Information retrieval today.* Arlington, Va: Information Resources Press.

Liu, J, Wu, Y & Zhou, L. 1999. A hybrid method for abstracting newspaper articles. *Journal of the American Society for Information Science,* 50(13):1234-1245.

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

O'Connor, BC. 1996. *Explorations in indexing and abstracting: pointing, virtue, and power.* Englewood, Colo: Libraries Unlimited.

Pinto, M & Cálvez, C. 1999. Paradigms for abstracting systems. *Journal of Information Science,* 25(5):365-380.

Pinto, M & Lancaster, FW. 1999. Abstracts and abstracting in knowledge discovery. *Library Trends,* 48(1):234-248.

Tenopir, C & Jascó, P. 1993. Quality of abstracts. *Online,* 17(3):44-55.

Wellish, HH. 2000. *Glossary of terminology in abstracting, classification, indexing and thesaurus construction.* Medford, NJ: Information Today.

Wheatley, A & Armstrong, CJ. 1997. Metadata, recall, and abstracts: can abstracts ever be reliable indicators of document value? *Aslib Proceedings,* 49(8):206-213.

14 THESAURUS CONSTRUCTION

Marlene Burger

Abstract

A thesaurus is a formally structured aid that is used to index and retrieve entities. Although thesauri are very useful, they also create problems that cannot always be solved. With the development of the Internet there has been renewed interest in thesauri. It is therefore necessary to understand the basic principles of thesaurus construction.

Introduction

Subject approaches in the electronic era have become a major way of finding information. Search engines are trying to fill the void on the Internet, yet users are becoming increasingly frustrated with the thousands of 'hits' from keyword searches (Taylor 2004:261):

> With the massive increase in availability of recorded information it becomes more and more evident that keyword searching alone will not suffice. Virtually every word in the English language has more than one meaning or sense, and many of those senses have more than one nuance; many words can be used as nouns, verbs, adjectives, or adverbs. Search systems that purport to allow the user to use "natural language" cannot yet successfully distinguish among meanings or parts of speech in very large general systems, although progress has been made in narrow subject areas.

We also know that authors writing about the same concepts often do not use the same words to express them, and users searching for the same concept do not think of the same words to search for it. The implication is that controlled vocabulary is needed to reconcile all the various possible words that can be used to express a concept and to differentiate among all the possible meanings that can be attached to certain words.

The purpose of this chapter is to introduce you to another aid to vocabulary control, namely thesauri. Thesauri are used in practice to retrieve information (e.g. from databases) or to index entities (e.g. periodical articles). You may even be required to compile a thesaurus for an internal database or for sources on the Intranet. Thesauri are also used in museums and archives. Information professionals may have doubts about the future role and value of thesauri; when compared to natural language or information organisation and retrieval systems (IOAR systems), there does not seem to be any need for thesauri, but with the Internet and information being freely available, more attention is being given to the compilation of thesauri for use on the Internet and to thesauri that can be used to retrieve information.

Aitchison, Gilchrist and Bawden (1997:xiii) explain it as follows: 'Other authors (with whom we agree) maintain that far from being dead, the thesaurus will continue to be a useful adjunct to retrieval systems, though increasingly hidden from the user … Certainly it is the case that around the world there are many systems that continue to use thesauri for controlled indexing, albeit many such systems offer free text searching as well.' Aids for vocabulary control solve some problems associated with natural language, but they also create problems of their own.

You will come across a variety of definitions of a thesaurus in the literature. It is generally accepted that a thesaurus is formally structured to show the relationships between terms, and it serves as an aid to indexers and others who use IOAR systems to retrieve information. A thesaurus is therefore an aid to be used when *organising* and *retrieving* information. Thesauri can be arranged alphabetically or hierarchically, they can be used for a computerised database or a printed index, and they are available in printed or electronic formats (online thesauri). The National Information Standards Organization (1994:1) explains a thesaurus as follows: 'A thesaurus, for the purpose of this standard, is a controlled vocabulary of terms in natural language that are designed for postcoordination. The need to control the information and use of terms stems naturally from the two basic features of natural language, namely synonyms (different terms representing the same concept) and polysemes or homographs (terms with the same spelling representing different concepts).'

Walker and Janes (1998:305) provide the following definition: 'An alphabetical listing of subject words that comprise the controlled vocabulary for a particular database. Terms from the thesaurus are selected by indexers to describe the document when it is being input, and by searchers to find documents for retrieval. Most thesauri also list subject-related terms for each entry as an aid to selecting the most specific terms.'

There is a different thesaurus for virtually every subject one can think of. The following figure (Figure 1) is an extended example (adapted from Lancaster & Warner 1993:8) of an IOAR system and the role of vocabulary. Note that there are a number of aids for vocabulary control, such as classification systems, thesauri and subject heading lists.

Characteristics of Thesauri

The following are characteristics of thesauri (more information on these characteristics later in the chapter):

- Are formally structured (e.g. alphabetically or hierarchically).
- Indicate a variety of indexing terms that the indexer/user might think of. Some of these terms are used as preferred terms in the thesaurus, while others are non-preferred terms or entry terms.
- Refer indexers/users from the terms not suitable for indexing (i.e. the entry terms) to the preferred terms used in the thesaurus.
- Indicate relationships with other terms. These relationships are indicated by means of special codes, such as BT (broader terms), NT (narrower terms), RT (related terms).

- Sometimes explain the meaning of indexing terms or how the thesaurus uses the term. This is generally indicated with the code SN (scope notes).

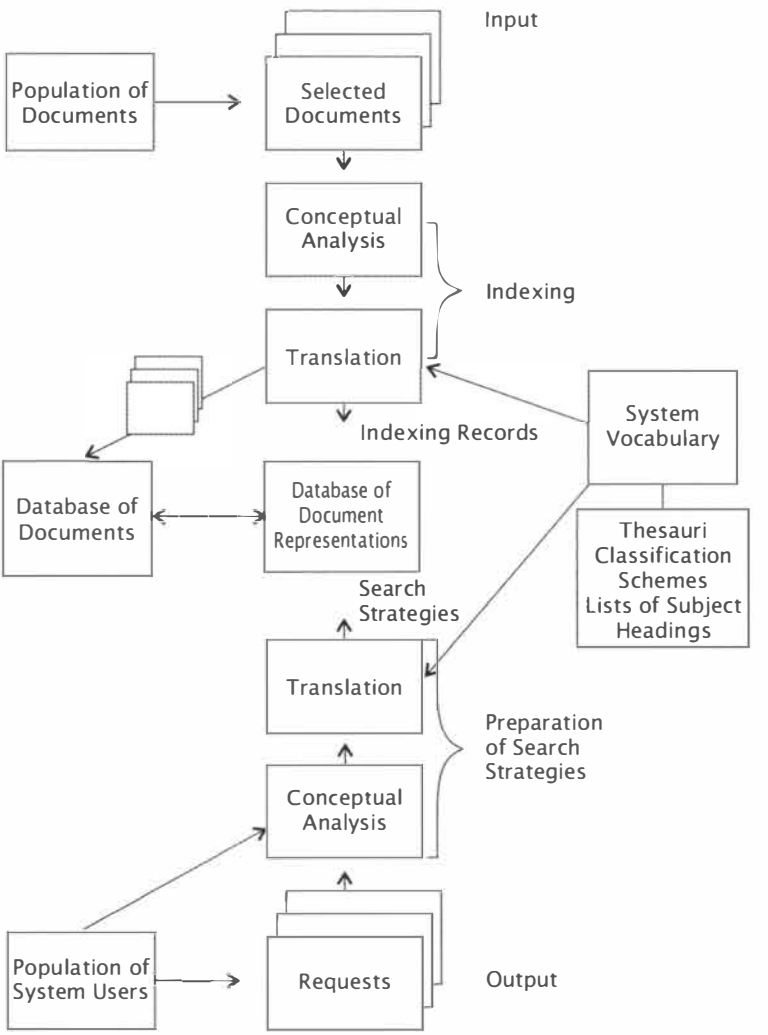


Figure 1: Example of an IOAR system

The following are examples (extracts) from different thesauri.

From the *Asis thesaurus of information science and librarianship* (1994:48):

ALPHABETICAL DISPLAY

library access
BT: access to resources
RT: information access
libraries
library acquisitions
USE: acquisitions (materials)
library administration
USE: library management
library assistants
USE: paraprofessional library personnel
library associations
BT: professional associations
RT: librarians
libraries
library automation
UF: electronic libraries
BT: automation
NT: computerized cataloguing
RT: integrated library systems
libraries
library buildings
BT: buildings
RT: libraries
library equipment
library collections
UF: book collections
book stock
holdings (library)
BT: collections
RT: library materials
library cooperation
USE: cooperation
library education
BT: education
RT: information science education
library equipment
BT: (hardware, equipment, and systems)
NT: library security systems
library shelving
RT: information technology
library buildings
library supplies
storage (materials)

From *Zoological record online* (an example of an online thesaurus) (Aitchison, Gilchrist & Bawden 1997:92):

Ref	Items	Type	RT	Index form
R1	24		9	* SEDATION
R2	0	S		Used only when technique itself is
R3	0	S		discussed/evaluated
R4	46	B	45	TECHNIQUES-
R5	0	N	7	ANAESTHESIA-
R6	0	N	6	DARTING-
R7	0	N	4	IMMOBILIZATION-
R8	0	N	6	TRANQUILIZATION-
R9	4	R	10	HANDLING-TECHNIQUES
R10	1	R	6	PHYSICAL RESTRAINING TECHNIQUES

In the thesaurus entry, the term entered is listed first. If scope notes are included, they are listed next, indicated by S under Type. Broader terms are listed under Type as B (for examples, TECHNIQUES-), narrower terms are listed under Type as N (DARTING-), and related terms as R (HANDLING-TECHNIQUES). At this point, users can select the appropriate term and search or explode the term. To explode, use the term, including hyphens, followed by an exclamation mark (!). This will search the term and all narrower terms listed under it.

From the *Thesaurus of ERIC descriptors* (1995) (some entries shortened):

Index terms
USE Subject Index Terms

INDEXES
SN Systematic guides to information, consisting of lists of logically arranged items with references that show where the items are located
UF Concordances
NT Citation Indexing
BT Reference Materials
RT Abstracts
Indexing
Information Retrieval
Subject Index Terms
Thesauri

INDEXING
SN Assignment of index terms to documents or objects in order to later retrieve or locate these documents or objects according to the selected concepts designated by the index terms
UF Subject Access
RT Abstracting
Indexes
Information Retrieval
Subject Index Terms
Thesauri

INFORMATION LITERACY
RT Library Instruction
Search Strategies
Users (Information)

INFORMATION RETRIEVAL
SN Techniques used to recover specific information from large quantities of stored data
NT Online Searching
RT Abstracting
Indexes
Indexing
Reference Materials
Search Strategies
Thesauri

INFORMATION SCIENCE
NT Library Science
RT Information Science Education
Information Scientists
Information Skills
Library Education

INFORMATION SCIENCE EDUCATION
SN Education concerned with the handling of information, comprising such information-oriented fields as Computer and Communication Science ...
BT Education
NT Library Education
RT Information Science

When Do We Use a Thesaurus?

A thesaurus can be used to

- assign indexing terms to entities
- determine appropriate search terms to satisfy an information need
- promote a general idea of a subject field
- offer a semantic map of the interrelationships between terms
- help provide definitions of terms

Generally we only use a thesaurus when searching for information; we call this type of thesaurus a search thesaurus. A search thesaurus can help a user to identify search terms and to think of more general or specific terms, synonyms and related terms. The purpose of a thesaurus is therefore to

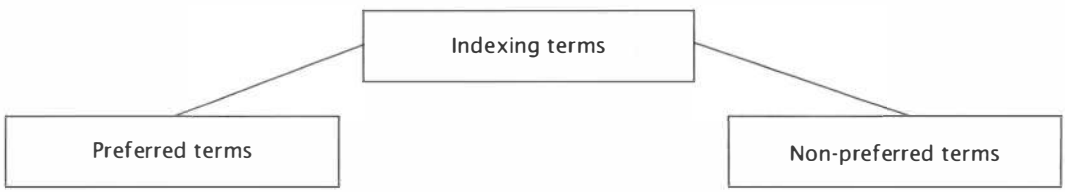
- indicate what terms may be used in an IOAR system

- indicate what terms may not be used in an IOAR system
- suggest terms to the indexer or person seeking information under which further cross-references can be made or information sought and thus promote consistency in indexing
- ensure that a topic is always described in the same way so that information is not lost because different words or phrases are used to describe the same topic
- indicate hierarchical and other relationships between terms

Structure of a Thesaurus

Thesaurus entries consist of indexing terms that include

- preferred terms
- forbidden or non-preferred terms, thus



A typical entry for a preferred term is as follows:

INDEXING TERM	
SN	(scope note)
UF	(used for)
BT	(broader term)
NT	(narrower term)
RT	(related term)
LIBRARIES	
SN	Institutions housing collections of systematically acquired and organised information resources, and usually providing assistance to users
UF	Resource centres
BT	Institutions
NT	Academic libraries
RT	Library funding

Not every indexing term refers to a broader, narrower or related term; it depends on the nature of the indexing term. Under each indexing term the indexer/user will be referred to other terms that are important. Some terms do not indicate relationships with other terms and these are called 'orphan terms'. A typical entry for a non-preferred term is as follows:

Entry term
U (use reference)
Resource centres
U **LIBRARIES**

The entry term directs the indexers/users from the terms not suitable for indexing to the terms preferred by the thesaurus (and also the IOAR system). Please note the use of font and case to differentiate between terms – that is bold, and capital and small letters.

Indexing Terms

Indexing terms are also referred to in the thesaurus as keywords, catchwords, descriptors and identifiers. An indexing term, which can comprise one or more words (try to avoid phrases), represents an idea or concept. Guidelines for formulating indexing terms are in a further section of this chapter.

Scope Notes

Indexing terms do not need to have a scope note. They are only used to explain or describe the meaning of a term if necessary. A scope note can be used

- to indicate the limitations of the meaning of a term

INCOME
SN Income of individual organisation or person. Otherwise use 'National Income'

- to indicate a series of subjects to be covered

MECHANICAL COMPONENTS
SN Includes manipulating parts, separating parts, linkages, internal parts, seals and gaskets, bearings, fasteners, chains, etc.

- to give instructions to indexers

INTERIOR LIGHTING
SN For lighting buildings or spaces, combine with appropriate terms, e.g. 'Shop lighting' use 'Shops' and 'Interior lighting'

Scope notes are sometimes used to explain terms that have the same spelling but different meanings. For these we use qualifiers, for example

cranes (birds)
cranes (lifting equipment)
mercury (metal)
Mercury (planet)
Mercury (Roman deity)

The scope note is not part of the indexing term. *A qualifier, however, is part of the indexing term.*

Relationships between Terms

We distinguish between three types of relationships between terms:

- equivalent relationships (USE and UF references are mostly used)
- hierarchical relationships (BT and NT are usually used to indicate broader and narrower terms respectively)
- non-hierarchical relationships (RT is generally used to indicate related terms)

Equivalent relationships **U, UF for 'use' and 'used for' are used to indicate equivalent relationships**	
True synonyms	**Examples**
Trade names	**COCA COLA** UF Coke Coke U **COCA COLA**
General use vs scientific use	Rabies or Hydrophobia
Local term vs international term	Aerial (UK) or Antenna (USA)
Competitive term for a new concept	Teleworking or Distance working
Standard terms vs slang	High Fidelity equipment or Hi-Fi equipment
Older vs newer terms	Wireless or Radio
Different stems	Terrestrial magnetism or Geomagnetism
Quasi-synonyms	**Examples**
Terms representing different poles of the same characteristic	STABILITY UF Instability Instability U **STABILITY**
Terms that overlap a great deal	Genetics / Heredity Accuracy / Precision
Specific concepts falling under one broad term, if the individual term is regarded as too specific	**CEREAL CROPS** UF Barley Rye Wheat
Various spellings of a term	**COLOUR** UF Color Color U **COLOUR**
Abbreviations	**VISUAL DISPLAY UNITS** UF VDU VDU U **VISUAL DISPLAY UNITS**

The person responsible for compiling the thesaurus will have to decide which terms will be the preferred terms and which the non-preferred terms. The thesaurus to be used by the indexer/user might look like this:

AERIAL
UF Antenna (when British terms are preferred)

Antenna
U **AERIAL**

COCA COLA
UF Coke (a choice must be made between the two terms)

Coke
U **COCA COLA**

Geomagnetism
U **TERRESTRIAL MAGNETISM**

Hydrophobia
U **RABIES**

RABIES
UF Hydrophobia (a choice must be made between the two terms)

RADIO
UF Wireless (because this is the older of the two terms)

TERRESTRIAL MAGNETISM
UF Geomagnetism (a choice must be made between the two terms)

Wireless
U **RADIO** (because this is the newer of the two terms)

The list of examples listed under quasi synonyms can be presented in a similar way. Further relationships between terms can be indicated when it is necessary. For example:

RABIES
UF Hydrophobia
BT Animal diseases
Diseases
RT Treatment

No further relationships must be indicated for entry terms:

Hydrophobia
U **RABIES**

We will now look at another type of relationship, namely hierarchical relationships. There are many more examples and fine divisions, but these examples should suffice for this introductory chapter on thesaurus construction. For more information, please consult the bibliography at the end of this chapter.

Hierarchical relationships **BT (broader or more general term) and NT (narrower or more specific term) are generally used to indicate hierarchical relationships**	
Genus/species relationships	**Examples**
This relationship is a link between the class or category and the members of the species. When one has to decide on narrower terms, always ask: Is it a type ...? If not, then it is a related term. For broader terms, always ask: Which is ... a type?	Dogs (genus) Poodles (species) Viral diseases Measles (species) Libraries (genus) Special libraries (species) **DOGS** **POODLES** NT Poodles BT Dogs
Part/whole relationships	**Examples**
System and organs of the body	Ear (whole) Middle ear (part) **EAR** **MIDDLE EAR** NT Middle ear BT Ear
Geographical location	Africa (whole) Southern Africa (part) Economy (whole) Commercial economy (part)
Hierarchical social structure	Methodist Church organisation (whole) Methodist district (part) Methodists (circuit) (part)
Instance relationships	Examples
This is the relationship between a general category of things and happenings, and individual examples	Sea (category) Baltic Sea (example) Black Sea (example) French cathedrals (category) Chartres Cathedral (example) Rouen Cathedral (example)
Poly-hierarchical relationships	**Examples**
Some terms occur in more than one hierarchy	Nurses (genus) Nursing administrators (species) Health administrators (species) Nursing administrators (species) **NURSING ADMINISTRATORS** BT Health administrators Nurses **ZAMBIA** BT Africa Southern Africa

For each of these examples, entries are made in the thesaurus and other relationships are indicated where necessary. We will now look at non-hierarchical relationships. There are many more examples and fine divisions, but these examples should suffice for this introductory chapter on thesaurus construction. For more information, please consult the bibliography at the end of this chapter.

Non-hierarchical relationships **RT (related term) is generally used to indicate non-hierarchical relationships. The purpose of related terms is to remind the user who is looking at one term, that he or she may also be interested in another order to retrieve relevant information**	
	Examples
Species of the same genus	SPECIAL LIBRARIES RT University libraries Both are species of the genus LIBRARIES
Object/Part	**BOOKS** RT Book covers
Object/Attribute	**INFORMATION SYSTEMS** RT Reaction time
Object/Process	**JOURNALS** RT Journal catalogues
Object/Application	COMPUTER DATABASES RT Information retrieval
Attribute/Process	**EFFECTIVITY** RT Effectivity measurement
Process/Object	**CATALOGUING** RT Catalogues
Process/Attribute	**INFORMATION RETRIEVAL** RT Precision

For each of these examples entries are made in the thesaurus and other relationships are indicated where necessary.

Compiling a Thesaurus

When constructing a thesaurus, there are a number of steps you should work through. These steps as well as the standards applying to thesaurus construction are discussed in the following paragraphs.

Step 1: Analyse why a Thesaurus is Necessary

You might need a thesaurus to index entities for an international database such as *ERIC, International information system for nuclear science*, or a local database about South African literature in library and information science, or to retrieve information from a database on the WWW. To do this, you need a sound knowledge of the IOAR system where the thesaurus will be used. It is also important to weigh up the merits of controlled vocabulary and natural language in the situation where the IOAR system will be used.

Step 2: Check if there Are Existing Thesauri that Can be Used or Adapted

Lists of thesauri are readily available. Aslib publishes an updated list every six months in the *Current Awareness Bulletin.* Some issues of *Knowledge Organization* also include lists. Some clearing houses collect copies of thesauri and they are also available online. To access these online lists, access the following website: http://www.fit.qut.edu.au/InfoSys/middle/cont voc.htm

- How much literature should be included in the IOAR system? This will determine the specificity of the indexing terms. More specific terms will have to be used if a large quantity of literature needs to be included.
- In which language should the thesaurus be compiled? (For example in English only, or in English, Afrikaans and Zulu? The latter is called a multilingual thesaurus.)
- Is the IOAR system manual or computerised?
- Who will use the IOAR system and thus the thesaurus (e.g. end-users or information specialists)? This will influence the choice of terms.
- What potential questions will be put to the IOAR system (for which queries will the thesaurus be used to identify the search terms)? When terms are collected for thesaurus entries, they must have literary and user warrant. Will the request be general or specific? If the request is specific, then the terms in the thesaurus must also be specific.

Step 3: Collect Terms to Include in the Thesaurus

The following sources can be used:

- Existing lists of terms (e.g. dictionaries and other thesauri).
- Subject literature. Terms appearing in the literature will be used by the IOAR system; these are known as terms with literary warrant. This approach is sometimes referred to as the empirical approach.

*Two methods can be used here: the **deductive** and **inductive** methods.*

With the deductive method terms are derived from entities and checked by experts. Broad classes are identified and then other terms are assigned to these classes on the basis of their logical relationships.

In the inductive method new terms are selected as they appear in entities for possible inclusion in the thesaurus. As each term is recorded, it is classified under a broad class. The thesaurus is constructed right from the beginning to have broader and narrower terms. Experts can be called in to help determine terms.

A combination of these methods can also be used.

- Subject experts. Experts can be approached to compile a list of terms or a committee can be established to decide on terms. These terms are known as terms with user warrant. This process is referred to as the committee approach or the bottom-to-top approach.

Step 4: Select and Formulate Indexing Terms

The thesaurus compiler(s) must decide which of the terms that have been identified should be included in the thesaurus. A decision must also be made regarding preferred terms and non-preferred terms.

- Verify that the terms selected are found in the entities and keep user preferences in mind (e.g. if a choice must be made between scientific and everyday terms).
- Consult relevant standards and guidelines when formulating terms.

Examples of important standards include (or any later editions):

British Standards Institution. 1987. *British standard guide to establishment and development of monolingual thesauri.* London: BSI (BS 5723).

International Organisation for Standardisation. 1985. *Guidelines for the establishment and development of monolingual thesauri. ISO 5964.* Geneva: ISO.

International Organisation for Standardisation. 1986. *Guidelines for the establishment and development of monolingual thesauri. ISO 2688.* 2nd edition. Geneva: ISO.

National Information Standards Organization. 1994. *Guidelines for the instruction, format and management of monolingual thesauri: an American National Standard.* Bethesda, Maryland.

- Use reference works such as dictionaries, encyclopaedias and other thesauri to standardise terms.
- Decide on the specificity of terms.
- Decide on the degree of precoordination necessary.

Step 5: Arrange the Terms in Categories so that They Are More or Less Related

Classification systems are particularly useful.

Step 6: Decide on the Structure of the Thesaurus (e.g. Alphabetical or Hierarchical)

If the thesaurus is to be organised alphabetically, use either a word-by-word or letter-by-letter arrangement:

- Word-by-word arrangement. Each word making up an indexing term is regarded separately. For example:

  Planet-gazing
  Planetariums
  Planetary orbits
  Planets

- Letter-by-letter arrangement. All spaces between words are ignored, as well as all characters (except for parenthesis), numbers and letters. Terms are filed according to the following sequence: left parenthesis, numbers 0-9, and letters A-Z. For example:
  Planetariums
  Planetary orbits
  Planet-gazing
  Planets

The thesaurus entries should be arranged in alphabetical order. Terms that are listed in each thesaurus entry should also be sub-arranged in alphabetical order. For example:

COOLERS
NT Air coolers
Beverage coolers
Milk coolers
Oil coolers

RT Air conditioning
Air conditioning equipment
Compressors

Step 7: Decide on the Format of the Entries

For example:

INDEXING TERM
SN (scope note)
UF (used for)
BT (broader term)
NT (narrower term)
RT (related term)

Step 8: Use a Standard Form or an Appropriate Thesaurus Program to Record Thesaurus Entries

- Decide which terms will be preferred terms and which non-preferred terms.
- Indicate relationships between terms.
- Add scope notes where necessary.

Step 9: Compile a Draft Thesaurus

Ask users and experts to test it for

- indexing of entities
- retrieving information from the IOAR system

Take note of the problems and revise the thesaurus as necessary.

Steps 10-14

Step 10: Write an introduction for the thesaurus and explain how it should be used.

Step 11: Edit the thesaurus.

Step 12: Copy the thesaurus and make it available online or in printed form.

Step 13: Send a copy to a clearing house.

Step 14: Maintain the thesaurus and evaluate its effectiveness on a regular basis.

Maintenance can be a time-consuming task. Terms have to be updated, new descriptions have to be added, existing entries have to be adapted and obsolete ones deleted.

Guidelines for Formulating Thesaurus Entries

The following basic guidelines have been taken from standards for thesauri and textbooks and include the many details associated with compiling a thesaurus.

Indexing Terms

- Noun and noun phrases (e.g. FEMALE WORKERS) are mostly used as indexing terms.
- Use adjectives as indexing terms in exceptional instances only. For example:
  miniature
  portable
  However, such terms should rather be used as compound terms in the form of phrases. For example:
  miniature portraits
  portable devices
- Avoid prepositions and verbs wherever possible. For example:
  book evaluation (not evaluation of books)
  communication (not communicate)

Singular versus Plural Form

- Nouns for tangible objects that can be counted (in the literature these are referred to as 'count' nouns) are used in the plural form, for example dogs, trees, children. The question 'how many?' is applicable to count nouns.
- For non-count nouns use the singular form, for example gold, silver, flour, sugar. The question 'How much?' is applicable to non-count nouns.

Be aware that there are exceptions to this rule, about which you will find more information in the literature.

Spelling

Standard spelling should be used. If a choice has to be made between English and American spelling, the user's preference should be kept in mind. There should always be a reference (U) from the term with the 'wrong spelling' to the term with the standard or preferred spelling. For example:

Archeology
U **ARCHAEOLOGY**
ARCHAEOLOGY
UF Archeology

Punctuation

Punctuation marks should be avoided if possible. Hyphens and apostrophes should not be included, for example

multicultural societies (not multi-cultural societies)
nonmetals (not non-metals)
postcoordinate indexing (not post-coordinate indexing)

The hyphen should be retained if its exclusion could lead to ambiguity or difficult reading. For example:

draw-off taps
n-types
parent-teacher relationships

Parentheses should be used if a qualifier for an indexing term is required. (A qualifier explains the meaning of a term and puts it into context so that the term is distinguishable from other terms with the same spelling but with different meanings.) For example:

Letters (alphabet)
Letters (correspondence)

Abbreviations

Try to avoid the use of abbreviations. They should only be used as preferred terms if they are very general and well known. For example:

North Atlantic Treaty Organization
U **NATO**

United States of America
U **USA**

Compound Terms

According to the *International ISO standard* it is preferable to use single terms for a compound concept, unless the compound concept would be clearer to the user. When

compiling a thesaurus it is often difficult to decide whether or not to use a compound term. The number of compound terms that occur in a thesaurus indicates the *level of precoordination* of the thesaurus. If many compound terms occur, the thesaurus will have a high level of precoordination. Examples of compound terms are

dried fruit
municipal engineering
philosophy of education
rubber legs
special libraries

When an indexing term consists of two words, preference should be given to the direct form of entry. For example:

mechanical engineering (not engineering, mechanical)
special libraries (not libraries, special)

When the selected terms have been arranged in alphabetical order, indicate their relationship with other terms. Make sure that for each term there is a *reciprocal entry*, for example

LIBRARIES
NT Special libraries

SPECIAL LIBRARIES
BT Libraries

Relationships Application

You have already been introduced to the relationships found in thesaurus entries. The following is a simple explanation of how these relationships and reciprocal entries work in practice.

Hierarchical Relationships

To determine hierarchical relationships, you should ask

Is … a type of …?

If the answer is YES, then is it a narrower term? For example:

Special libraries is a **type** of library, therefore

LIBRARIES
NT Special libraries

In the thesaurus a specific entry is made for special libraries: Libraries is then referred to as the broader term. Under both Libraries and Special libraries a number of relationships can be indicated, for example

LIBRARIES
NT Academic libraries
Public libraries
Special libraries
RT Library buildings
Library personnel

SPECIAL LIBRARIES
BT Libraries
RT Actuality services
Information services

For the terms Academic libraries and Public libraries, similar entries are made, and each time Libraries is used as the broader term.

Non-hierarchical Relationships

To determine non-hierarchical relationships, two questions must be asked:

- Is ... the same as ...?
- Is ... a type of ...?

If the answer is NO to both questions, then you must think of possible non-hierarchical relationships. Non-hierarchical relationships help the user and indexer to think of other terms which might be useful. If, for example, you are looking for information about **Subject analysis**, think of the following: Would information about **Conceptual analysis** be useful? Would information about **Indexing** be useful? Would information about **Abstracting** be useful? You can continue to determine related terms in this way. Related terms are indicated as follows:

SUBJECT ANALYSIS
RT Abstracting
Conceptual analysis
Indexing

A thesaurus entry can also be made for abstracting, with Subject analysis as a related term:

ABSTRACTING
RT Subject analysis

Further relationships can be indicated under both Subject analysis and Abstracting. For example:

ABSTRACTING
RT Abstracts
Indexing
Subject analysis

SUBJECT ANALYSIS
RT Abstracting
Conceptual analysis
Indexing

Similar entries can be made for Indexing and Conceptual analysis – each time they must refer to Subject analysis as a related term.

Equivalent Relationships

To determine equivalent relationships, one should ask

- Is … the same as or almost the same as …?
- Do I want to use … instead of …?

If the answer is YES to any of these questions, then you must think of possible equivalent relationships. Equivalent relationships are used to guide the user of the thesaurus (indexer or enquirer) away from the terms not regarded by the system as suitable for indexing, to the preferred terms with which the concept can be described. To indicate these relationships, cross-references are used (U, UF). For example:

Librarianship
U **LIBRARY SCIENCE**

Library employees
U **LIBRARY PERSONNEL**

LIBRARY PERSONNEL
UF Library employees

LIBRARY SCIENCE
UF Librarianship

Draw up the thesaurus in typed form – use an existing thesaurus as an example. Keep the thesaurus up to date by making changes, adding new terms and dropping or replacing obsolete terms.

Thesaurus Software

Take note of available thesaurus software and the advantages this software offers. If you would like to experiment with such programs, you can go to the following websites. Some of them offer demonstration programs that you can download:

LEXICO: http://www.pmei.com/lexico.htm

MULTITES: http://www.multites.com

QUESTANS LIMITED: http://www.questans.co.uk
The Hierarch Thesaurus Manager: http://www.ozemail.com/au%7Esisnsw/hierarch.htm

There are special programs available for constructing thesauri. These programs make provision for the following:

- typography and sorting (e.g. correct use of capital and small letters and also special characters)

- displaying the record for each entry (this includes the relationships between terms)
- distinction between non-preferred and recommended terms

When evaluating such a program for possible use, ensure that the program

- does not limit the number of characters in a term, the levels of hierarchy, or the number of relationships that may be indicated
- is able to handle searches on full terms (e.g. libraries) and also truncated terms (e.g. librar…)
- shows entries in such a way that a term can be displayed with all its relationships
- makes it easy to create and edit terms
- supports the tracing of errors (e.g. duplicate entries and inconsistent application of relationships)
- makes cross-references (relationships should be correctly indicated and terms to which there are cross-references should actually exist)
- is able to delete terms
- is able to add candidate terms provisionally
- allows other fields (e.g. classification notations) to be included
- provides printed reports from the thesaurus in a format required by the user
- is able to copy a sub-collection or a micro-thesaurus of thesaurus entries from one thesaurus to another

BIBLIOGRAPHY

Asis thesaurus of information science and librarianship. 1994. Medford, NJ: Learned Information.

Batty, CD. 1989. Thesaurus construction and maintenance survival kit. *Database*, 12(1):13-20.

Aitchison, J, Gilchrist, A & Bawden, D. 1997. *Thesaurus construction and use: a practical manual.* 3rd ed. London: Aslib.

Borko, H & Bernier, CL. 1978. *Indexing concepts and methods*. New York: Academic Press.

Lancaster, FW. 1986. *Vocabulary control for information retrieval.* 2nd ed. Arlington, Va: Information Sources Press.

Lancaster, FW & Warner, A. 1993. *Information retrieval today*. Arlington, Va: Information Resources.

National Information Standards Organization. 1994. *Guidelines for the instruction, format and management of monolingual thesauri: an American National Standard*. Bethesda, Maryland.

Orna, E. 1983. *Build yourself a thesaurus: a step-by-step guide.* Norwich: Running Angel.

Ovens, CS. 1989. Methodology for constructing a thesaurus in the linguistics and literature fields. *South African Journal of Library and Information Science,* 57(1):45-50.

Taylor, AG. 2004. *The organization of information.* 2nd ed. Westport, Conn: Libraries Unlimited.

Thesaurus of ERIC descriptors. 1995. Edited by JE Housten. 13th ed. Phoenix, Ariz: Oryx Press.

Walker, G & Janes, J. 1998. *Online retrieval: a dialogue of theory and practice*. Englewood, Colo: Libraries Unlimited.

15 TECHNICAL PRESENTATION OF BOOK INDEXES

Umashanie Reddy and Omesh Jagarnath

Abstract

An index is the soul of a book, and if a book is worth publishing at all, it is worth an index (Hewitt 2003:201). According to Cleveland and Cleveland (1990:100) good indexing is not a casual job but the result of a professional activity carried out by people with proper training and experience. Procedures and techniques acquired over the years have to be learnt and followed. The purpose of this chapter is to familiarise the reader with the technical presentation and editing procedures of a book index. Many of the issues in this chapter have been covered in other chapters as well, but since this is such an important part of the indexing process, this chapter provides further examples and explanations. The general principles can also be applied to periodical indexes and thesauri.

Introduction

The editing process of indexes cannot be clearly distinguished from the selection and formulation of entries. Most indexers edit briefly as they work, making sure that each new entry is free of errors and that it conforms to existing entries. Others prefer to input all their raw entries as fast as possible without any checking or revision. On completion, the index is scrutinised in its entirety to standardise the headings, add reversed or permuted headings, construct subheadings, add cross-references, ensure consistency of style, and the like. Others prefer to proceed more slowly and control the structure of the growing index. For example, when subheadings become necessary (due to the magnitude of locators, or because several different aspects of the subject are identified) the locators are grouped accordingly by appending suitable words or phrases; anomalies in names and variations in spelling and capitalisation are checked when they arise; alternative headings (reversed or permuted) and cross-references are created when the need arises. *No kind of preliminary editing cancels out the need for final editing of the index.* It is important to remember that perfection is often not achievable in the formulation of index entries – to observe consistency in one feature can cause inconsistency with regard to another, and to follow a technically correct guideline can result in something that looks odd in a particular index. A pragmatic approach, treating the text with respect and making access as simple as possible for the user, is the best (Booth 2001:127-128).

Editing and Presentation

The finished index must provide a balanced and relevant coverage of the entity's significant content; it must enable access by all likely terms, be clearly laid out, be preceded by an introductory note (if necessary) and meet the specifications agreed upon by the publisher. *Editing ensures consistency of form and structure. Presentation influences the readability and use of the index.* Users must find the terms they are looking for with ease and understand them unambiguously, as well as note relevant references. Presentational features are normally already in the indexer's mind during input and some can be incorporated from the start, for example capitalisation, punctuation, indication of major references and illustrations. The final presentation, spacing and layout, additional styling and preparation for the chosen form of copy (paper, disk, camera-ready) can only be given attention once the individual entries and the overall structure is in place (Booth 2001:130).

Whether the index is edited on screen, or on a print-out from which the corrections are then keyed in, depends on the indexer's preferred method of working. Working on a print-out takes extra time, but it is easier to scrutinise a print-out, page back and forth to compare entries, see where improvements and corrections are needed, make notes and keep track of amendments. The last opportunity for editing an index for a printed entity is during the checking of the index page proofs.

A poor index is difficult to repair without an extensive makeover. At this stage the focus should be on such things as misspelling, large blocks of undifferentiated locators, too many subheadings under a heading, related information scattered throughout the index without cross-references. Dead-end cross-references should be corrected and page reference accuracy should be randomly checked. According to Cleveland and Cleveland (2001:190-191), the tasks can be summarised as follows:

- Correct alphabetising.
- Divide long multi-level headings.
- Eliminate synonymous headings by consolidation under preferred terms.
- Recheck the editing.
- Correct too many locators attached to an entry.
- Correct ambiguous headings.
- Verify cross-references.
- Correct spelling and capitalisation.
- Check punctuation and capitalisation.
- Check need to add entries.
- Ask: Are main headings relevant and/or needed?
- Ask: Are locators correct?
- Ask: Is the index overdone?
- Ask: Is the index underdone?
- Bottom line: Does the index work?

Factors Influencing Editing

Booth (2001:132-135) explains 11 factors influencing the amount of editing to be done. In this chapter, these are only mentioned briefly since they may differ from one index/indexer to another, but should be attended to:

- Quality of the input. As already mentioned, it depends on the indexer's preference when to edit (while indexing or after the index is completed). If more time is spent during the input stage, less time will be required for editing at the end.
- Length and complexity of entries. An index containing many multiple-word entries, or entries with several subheadings or sub-subheadings, or having complicated locators, will need more editing time.
- Number of problems noted during input. Indexers often note errors and variations in spelling in the text. Rather than spending time on each one as they arise by consulting the author, editor or publisher, it is better to deal with them together during editing.
- Incidence of special characters in entries. These include mathematical symbols, accented letters, other alphabets, and so on, and may not be available on the indexer's keyboard; or if available, they may be reproduced differently when transferred to another system (e.g. at the publisher). Special characters not available on the keyboard can be marked during input by using other characters so that they can be converted or noted during editing. This will need more editing time.
- Number of entries. The longer the index, the longer it takes to edit, and the more difficult it is to spot inconsistencies between entries.
- Mismatch between the length of the index and the space available for it. Even though the indexer knows from the start how much space is available for the index, it may still be too long. If it is not possible to make more space available by using a smaller font than originally agreed upon, the indexer is faced with the difficult and time-consuming task of reducing or removing entries. It may also happen that the index is too short for the allocated space.
- Style requirements. Most features like capitalisation, punctuation, use of italic or bold typeface can be done during the input stage. Others must be left until the final editing, for example putting locators for major references in bold type.
- Degree of autonomy of the indexer. Usually the indexer is entirely responsible for the content of the index – the publisher or editor only ensures that the submitted index meets the agreed specifications. Sometimes the index needs to be approved by the editor, publisher or author of the text. If, for example, suggestions are made for the addition and/or removal of entries, or changes of wording, the index should be referred back to the indexer.
- Time available. It can happen that the deadline for the index does not allow for thorough editing. In such cases it is better to edit the index while working on it. The indexer can try to negotiate for more time, but it is not often granted.
- Producing the index. The physical form in which the index is prepared varies according to the requirements of the publisher: e-mail, on disk, hard copy, et cetera. A hard copy

gives a clear view of the index and its layout and allows for notes to the publisher regarding special features. Be aware that the page dimensions, the font and size used in a typeset document are almost always different from those used by the indexer. Pay particular attention to subheading sequences that may be broken by ends of columns or pages. This is especially important when a camera-ready copy is required (in instances where the indexer has the suitable software). In-house writers producing documents on desktop publishing systems may compile their own indexes using embedded codes placed in the text. For more information on this, see chapter 8 on embedded indexing.

- Proofreading. This takes place after the index has been printed. It should be done by the indexer, but is often dealt with by the editorial staff of the publisher.

At all times the indexer should ensure familiarity with the agreement and adhere to its stipulations regarding the technical presentation of the index. Should problems arise, contact the editor or publisher and discuss the alternatives. The following paragraphs cover specific issues – some issues can be dealt with while constructing the index, while others should be left until the final editing.

Specific Tasks

Check for Errors

Activate a spell check during the indexing process. Typing and spelling errors will thus be indicated. If indexing on cards, each entry must be checked for correct spelling. Ideally this should be done by another person (time permitting). It is imperative in book or periodical indexing to adhere to the terminology used by the author. One's own preferences cannot substitute those of the author, for example 'archeology' for 'archaeology' or 'organization' for 'organisation'. When compiling a thesaurus, system preferences will have to be followed. Some indexers prefer to do a character-by-character (instead of a word-by-word) reading of the index. With a word-by-word reading the visual impression of some indexing terms might 'look' correct, for example 'geranium' instead of 'germanium' or 'Van der Berg' instead of 'Van den Berg'. If the numeral '0' and the letter 'O' are incorrectly keyed in, a computer-sorted index will misplace the entry.

Check for Scattering

Make sure that a single topic is not scattered under different word forms, synonyms, or near-synonym words in the index, for example 'goose' and 'geese', 'dexterity' and 'dexterous', or 'civil law' and 'common law'. If terms are left scattered under more than one heading, it may confuse users. Rather use just one form for each topic (e.g. plural, nouns). Some instances may require cross-references.

Check Multi-word Headings

These are also known as compound terms or complex terms, and can also be phrases, that is consisting of more than one word. Some examples are 'descriptive cataloguing',

'computer-assisted indexing' and 'progressive subheading order'. Check such headings for possible additional entries, for example 'cataloguing', 'indexing' and 'subheading order'. Under these broader terms the specific terms may be listed, or cross-references may be used, depending on the structure of the index. At this stage, also check too long headings – they can be reduced by the removal of prepositions for example, or perhaps rephrased.

Check Cross-references

Cross-references can be added while indexing, but at the final editing they should be checked in the context of the full index. Some may be discarded or others may be added. All cross-references must lead to headings that exist in the index and that are followed by locators, otherwise dead-end cross-references will ensue. Cross-references should be used only when necessary. Some headings may require more than one cross-reference. Examples of cross-references:

cataloguing *see also* subject cataloguing
dedicated indexing software 207-209
embedded indexing 38, 121-127, 206
  editing 130-133
  software 210-215
indexing
  as classification 15-19
  defined 5-6
  policies 22-32, 45
indexing software *see* dedicated indexing software
  embedded indexing 210-215
subject cataloguing 15-19
subject indexing *see* indexing

Check Locators

The accuracy of locators (e.g. page numbers, volume and part numbers, columns) is extremely important since it affects the access to the information contained in the entity. Errors (including possible typing errors) should be constantly checked to ensure accuracy. Page ranges must be consistent, for example 331-337, or 331-37, or 331-7 (full ranges are often preferred to reduced forms). Long strings of locators (i.e. without subheadings to indicate parts or points) are mostly irritating to the user since he or she must work through all locators to find a needed piece of information.

For example:

metadata 14, 69, 91-92fig, 123-125, 139, 141, 143-144, 147-148,149, 151-152, 153-154, 155, 159, 164, 190-192, 202, 203, 335-341

There are different ways of solving this problem, of which only the most frequently used two are illustrated. Bold type can be used to highlight locators leading to major information or emphasising a particular point:

metadata 14, 69, 91-92fig, 123-125, 139, 141, 143-144, 147-148,149, **151-152**, 153-154, 155, 159, 164, 190-192, 202, 203, **335-341**

A more sensible solution is to use subheadings:

metadata
  access to 149, 202
  administrative 147-148
  arrangement 335-341
  characteristics 143-144
  content 69
  creation 159, 190-192
  definition 139
  design 143
  granularity 141
  implementation 151-152
  maintenance 91-92fig
  mapping between formats 155
  preservation 148-149
  registries 153-154
  searching 123-125
  standards for 164
  structure 142, 150-151
  systems 140
  technical 150-151
  types of 141, 152, 203
  use of 14

All the locators should be represented by subheadings – there must be no leftover locators. However, sometimes long strings of locators are unavoidable, for example the listing of locators where an author is cited in the text of an entity:

Lancaster, FW 3, 15, 26, 34, 40, 61, 68-69, 79, 84, 89, 92, 101, 113, 126, 148, 184, 188-189, 236-237, 239, 244-245, 292, 296, 320-321, 383-384, 396

Check for Qualifiers

Some headings may need a qualifier to make its meaning clear, for example when a word has more than one meaning or if persons share the same name:

navigation (aeronautics)

navigation (astronautics)

Barnard, Chris (author)

Barnard, Chris (surgeon)

Check Headings Beginning with the Same Word

Headings beginning with the same word should be arranged in such a way that there is no unwanted separation of terms or odd-looking subheadings. Therefore:

mass-energy equivalence
Mass, Jason
mass media
mass media music
mass (substance)
Mass Transfer Union
mass transportation

is preferable to

Mass
-energy equivalence
Jason
media
media music
(substance)
Transfer Union
transportation

Check Alphabetical Arrangement

Preference must be given either to word-by-word arrangement or letter-by-letter arrangement; subheadings must follow the same arrangement as the main headings; take into account punctuation, prepositions, articles, et cetera; look at chronological sequences; hierarchies, progressive arrangement, and the like.

Word-by-word and Letter-by-letter Arrangement

In the word-by-word method, spaces between words have value so that shorter words come before longer ones, for example 'snake charmers' will be filed before 'snakebites'. Hyphens (and dashes and diagonal back or forward slashes) are treated the same way as spaces. For example:

snake charmers
snake dances
snake-eels
snake-flies
snake plants
Snake River
snake venom
snakebites
snakeheads (Fishes)
snakes

In the letter-by-letter method, spaces and other kinds of characters are ignored. For example:

snakebites
snake charmers
snake dances
snake-eels
snake-flies
snakeheads (Fishes)
snake plants
Snake River
snakes
snake venom

The other kinds of arrangement mentioned at the beginning of this section (e.g. hierarchical, chronological, progressive) are not discussed since they appear only in certain types of index, for example in indexes to history books, or in manuals where technical processes or developmental stages are discussed. Also pay attention to initial articles, abbreviations, prefixes, numerals, broken order, diacritics and special characters.

Layout of Entries

Subheadings should be clearly distinguishable from the headings to which they are subordinate, as well as from any sub-subheadings that may follow. Two methods are used: setting out and running on. The set-out method gives a clearer display and can accommodate many levels of subheading, while the run-on method takes up less space, but is limited in the number of subheading levels it can handle. In the set-out layout, each subheading (wherever there is more than one) starts on a new line, with indentation used to show subordination – any further levels of subheading are progressively and evenly indented. In the run-on layout subheadings continue on the same line, with punctuation indicating subordination and when the line is full, it turns over to the next line that is indented by one space. Only two levels of subheading can be accommodated if the subheadings are run on directly after the heading. A combination of the two methods is also possible (Booth 2001; Mulvany 1994).

The following examples illustrate the two layouts:

Run-on layout:
Internet: communication 10-18; downloading 317-319; networks 5; search engines 215-217 (*see also* searches); tools 20
Set-out layout:
Internet
communication 10-18
downloading 317-319
networks 5
search engines 215-217 *see also* searches
tools 20

Spaces and Typestyle

Spaces between the letters of the alphabet in the index makes it clear where a new letter starts and ease the use of the index. For example:

Sears list of subject headings
subject analysis
systems

templates
terminology

Unger, Richard
Unicode
user behaviour

If space is available, a letter can be placed at the beginning of each letter sequence. For example:

S
Sears list of subject headings
subject analysis
systems

T
templates
terminology

U
Unger, Richard
Unicode
user behaviour

Please note that it is common practice to present titles of works and terms in foreign languages (such as Latin and Greek) in italics. Also ensure that the use of punctuation between page numbers and the use of brackets for qualifiers is consistent.

Conclusion

This chapter has outlined basic principles on the technical features of presenting a book index. However, some issues were not covered in this chapter. The best source to consult for guidelines is a standard such as the BS ISO 999 (see the bibliography that follows) and/or a reputable indexing manual (see the bibliography). Should the appearance of the published version of the index be unsatisfactory when received at the page-proof stage, the publisher should be consulted.

BIBLIOGRAPHY

Booth, PF. 2001. *Indexing: the manual of good practice.* München: Saur.

Borko, H & Bernier, CL. 1978. *Indexing concepts and methods.* New York: Academic Press.

British Standards Institution. 1996. *Information and documentation – guidelines for the content, organization and presentation of indexes.* BS ISO 999:1996. London: British Standards Institution.

Butcher, J. 1992. *Copy-editing: the Cambridge handbook for editors, authors and publishers.* 3rd ed. Cambridge: Cambridge University Press.

Cleveland, DB & Cleveland, AD 1990. *Introduction to indexing and abstracting.* 2nd ed. Englewood, Colo: Libraries Unlimited.

Cleveland, DB & Cleveland, AD. 2001. *Introduction to indexing and abstracting.* 3rd ed. Englewood, Colo: Libraries Unlimited.

Gregory, GM. 2000. Current trends in book indexing. *Information Today,* 17(5):69.

Hewitt, AR. 2003. Another 'How to make an index'. *The Indexer,* 23(4):200-201.

Lancaster, FW. 1998. *Indexing and abstracting in theory and practice.* 2nd ed. Champaign, Ill: University of Illinois, Graduate School of Library and Information Science.

Lipetz, B-A. 1989. The usefulness of indexes, in *Indexing: The state of our knowledge and the state of our ignorance: Proceedings of the 20th Annual Meeting of the American Society of Indexers*, edited by BH Weinberg. Medford, NJ: Learned Information:111-120.

Mulvany, NC. 1994. *Indexing books.* Chicago: University of Chicago Press.

University of Chicago Press. 1993. *The Chicago manual of style.* 14th ed. Chicago: University of Chicago Press.

Wellisch, H. 1995. *Indexing from A to Z.* 2nd ed. New York: Wilson.

16 STANDARDISATION

Peter Duncan

Abstract

International standards for all types of indexing are considered, especially through the ISO, and information is given on where to find these documents. A significant one, ISO 999:1900, on the content and presentation of indexes is summarised. This also discusses the quality of indexes, how headings should be presented and chosen and the form of the final document for the printers. A useful document, ISO 5963:1985, on the method of indexing designed to facilitate communication between all people involved in the indexing process, is described. Reference is also made to other relevant documents.

Introduction

The International Standards Organization (ISO) has set out documents that give the standards of practice and technique to be applied by indexers internationally, with the aim of achieving high quality indexes. The standards are designed for indexes of documents, their contents and of non-print items, media and three-dimensional objects.

South Africa has no national policy, so one could adapt the ISO standards for local use. Quality publishers will base their house style guidelines on the ISO norms. The ISO documents can be obtained in PDF or printed format through the ISO website, http://www.iso.org/iso/en/CatalogueDetailPage.

They are detailed, designed to cover every contingency, and pricey.

Some Relevant ISO Documents
http://www.iso.org/iso/en/CatalogueDetailPage

The dates after the document number refer to the latest revision as at May 2004 /H2/ *ISO 999:1996:information and documentation.* This gives 'guidelines for the content, arrangement and presentation of indexes to books, periodicals … but also non print materials, such as electronic documents, films, sound and video recordings, maps and three-dimensional objects' (http://www.iso.org/iso/en/CatalogueDetailPage).

Summary

The document opens with a reference to other relevant ISO documents, especially ISO 5963:1985 documentation. It suggests the function of an index as well as quality controls.

These should meet the needs of those seeking information and the indexer should be impartial when choosing subject matter and terminology. An index should be sufficiently detailed and consistent. Authority files and thesauri will help achieve this aim. The document should be discussed with the author and publisher. The expected needs of the user should be decisive if there are conflicts between authors, publishers, database managers and indexers. Indexers need access to the entire document, for example to the illustrations, captions, maps and charts. The final index should be proofread by the indexer. Indexers should have the opportunity of being named in the publication in acknowledgment of their work.

Content and Construction

The entire document should be covered and any parts of the text excluded should be pointed out in an introductory note to the index. An introductory note can be included, where necessary, to explain how the index has been constructed. One single index is ideal, but sometimes separate indexes might be needed, for example for the indexing of chemical formulae or patent numbers. These separate indexes should be clearly marked.

Headings

Headings and concepts should be chosen from the document and influenced by the purpose of the document. One term should be used always for the same concept, with the appropriate *see* cross-references. This also applies to spelling variations and abbreviations. Linked terms that need to be indexed together will also need references, for example

evil *see* good and evil

Headings should consist of nouns, only modified by adjectives where necessary, for example

cutting tools

The plural should be used for countables, for example

animals

When the plural and singular have different meanings both can be used, for example

building
buildings

The spelling conventions of the text, or authority file should be followed. Double-word concepts should not be inverted or subsumed, for example

balance sheet *and not*
sheet, balance

Prepositions need only be used to avoid ambiguity, for example

computers for management
computers, management of

Personal names should be given as fully as possible. Variations should be cross-referenced and compound names should be entered under the first part, for example

Xavier Pérez de Cuéllar, Javier

Over time, preferred terms change. In serial publications or newspapers one might have to note, for example, that before 1950 the word 'wireless' was used and not 'radio'. These changes will need to be explained in the index. This also applies to names, for example

Burkina Faso *see also* Upper Volta for references before November 1984

This can also happen in books. The indexer should add in parenthesis 'now [new name]' with an appropriate cross-reference, for example

Lobel, Alice (now Alice Synkova) *and*
Synkova, Alice *see* Lobel, Alice

Corporate bodies should be entered as fully as needed to distinguish them from those with similar names. Both personal and corporate names should be used in the most recent commonly used form. Geographical names should also be as full as necessary for clarity. Titles of documents should be distinguished by underlining or italicised. If necessary, qualifiers can be added in parenthesis, for example

Ave Maria (Gounod)
Ave Maria (Schubert)

Locators

The document discusses locators, that is page numbers, volume numbers, dates, et cetera. These will depend on the material being indexed. A single-volume work will generally need only a reference to page numbers, but an index to a collection of documents will need, at least, reference to the author and title of the document and its pagination and/or volume number.

Arrangement of Entries

Filing rules or the arrangement of entries is discussed. Ideally, upper and lower case letters should be given equal value. The force of modified characters, such as ä, å and õ, should follow local practice. The value of a blank space depends on whether letter-by-letter or word-by-word indexing is being followed. Headings consisting of more than one word should be arranged word by word, in which a space files before a letter; for example 'New York' files before 'Newark'. The alternative letter by letter may be used if needed; then 'Newark' will file before 'New York'. Arabic and Roman numerals should file before the alphabetical entries, but according to their numerical value. For example:

1:30 am
XX century cyclopedia and atlas
1001 nights
1066 and all that

They can be filed as if they were spelt-out words.

The filing position of any index heading must be clear and predictable. The alphanumeric arrangement is the most practical. Other arrangements, such as patent numbers, may be needed in technical indexes. Subheadings should follow the same pattern as headings. This can be modified chronologically or numerically if the text suggests it.

Presentation of Printed Indexes

The presentation of the final copy for processing should be done in a way that meets the publisher's (or corporate body's/author's) needs. This might be typescript or machine-readable format accompanied by hard copy for individual documents. For collections of documents it might be anything from card indexes to electronic publications.

The copy should give the typesetter clear indications of the desired layout. The conventions of typesetter's indicators, of the country in which the typesetting is to be done, should be followed. There should be a blank line between each alphabetical sequence. Each page on which the index is printed, should have a running headline. This should inform users where they are in the index and, in the case of multiple indexes, which one they are using – such as the author index as opposed to the geographical one. To indicate the scope of a page, the first and last index entries can be used. If an index entry runs onto a new page the heading and maybe the subheading should be repeated, followed by the word 'continued' in parenthesis. The index is usually presented in two columns per page.

The hierarchy of headings can be indicated by either the set-out layout or the run-on layout. With the set-out style, hierarchy is indicated by spacing and indentation, for example

```
Aristotle
    debt to Plato 23, 46;
    literary criticism in 35, 74, 89-93
      on Aeschylus 101-104, 279
      on Aristophanes 195
      on Euripides 104-126, 187
```

The run-on style gives hierarchy by means of punctuation between entries, for example

```
Aristotle
  debt to Plato 23, 46; literary criticism in 35, 74, 89-93; on Aeschylus 101-104; on
  Aristophanes 195; on Euripides 104-126, 186
```

Addendum

This gives references to other relevant ISO documents, such as ISO 9:1995 *Documentation – transliteration of Slavic Cyrillic characters into Latin characters,* ISO 690:1987 *Documentation – bibliographic references – content form and structure,* and ISO

5776:1983 *Graphic technology – symbols for text correction.* These can all be found via the web page mentioned earlier. Multilingual indexing is prevalent in Southern Africa, and so the following document may be of interest: ISO 5963:1985 *Documentation – guidelines for the establishment and development of multilingual thesauri.* This needs to be used in conjunction with ISO 2788:1986 *Documentation – guidelines for the establishment and development of monolingual thesauri.* One of these will now be considered:

/H2/***ISO 5963:1985***

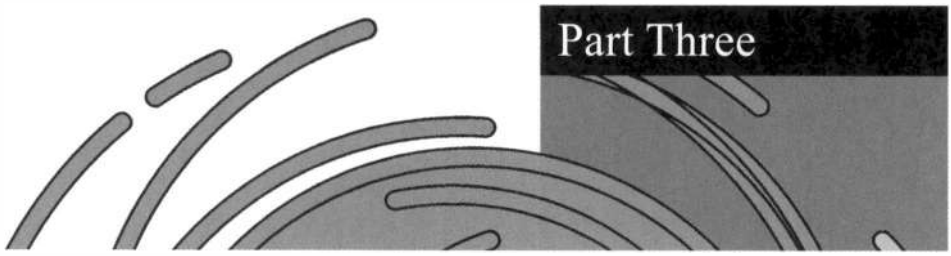
Documentation – methods for examining documents, determining subjects, and selecting index terms.

Abstract

Describes general techniques for document analysis whish should apply in all indexing situations. These techniques can be employed by any agency in which human indexers analyse the subjects of documents and express these documents in indexing terms. It is intended to promote standard practice within an agency or a network of agencies as well as between different indexing agencies, especially those which exchange bibliographic records (http://www.iso.ch/iso/en/CataloqueDetailPaige).

Conclusion

The ISO standards which have been discussed are flexible, since local, national and publisher's various needs are taken into account. They are sound guidelines for any indexer to follow.



SPECIAL FIELDS

17 AFRICANA

Yvonne Garson, Peter Duncan and Christie Theron

Abstract

The concept of Africana is extensively analysed with a focus on the problems experienced in indexing this genre. The history and importance of various Africana indexing tools is given. The pre-Internet tools SABIB, SANB and ISAP are discussed as well as the unpublished Fairbridge Index. The background history to SABIB is given, with the pioneering work of George McCall Theal and Sidney Mendelssohn. The value of the work is explained. The reader is introduced to their online successors and the basic new Internet tools African Studies and South African Studies. Into the second decade of South Africa's democracy an important new resource, the DISA database on the freedom struggle, is mentioned. These can be useful for subject headings and name authorities in indexing.

Introduction

Africana is a well-known and often used word or concept in South Africa and elsewhere. It is a word used by booksellers, librarians, book collectors and dealers in antique maps, furniture, et cetera. But what does this word or concept really mean?

Die handwoordeboek van die Afrikaanse taal (HAT) describes Africana as books, documents, maps and so on, of historical importance with regard to Africa [translation].

It is interesting to compare this with a definition of Americana from *Funk and Wagnall's* (1968): 'Material relating to the Americas, whether [it is] printed about or in, the Americas, or written by Americans.'

Perhaps we should first of all provide a better definition of the concept 'Africana'. Africana has to do with Africa. The suffix *-ana* is taken from the Latin, and refers to what belongs to, or has to do with, a certain place. It can refer to both a person or a place, and may include sayings, anecdotes or publications about that person or place. Africana literally means material dealing with Africa.

The view that all publications dealing with Africa are regarded as Africana is an interpretation held by many countries, although, according to Bradlow (1970b), this was the prevalent view in South Africa too, until about 1939. Thereafter, the Zambezi was regarded by South Africans as being their northernmost border. Currently, the Limpopo serves this purpose. South Africa (as well as Lesotho, Botswana, Swaziland and Namibia) is today regarded as Africana terrain. This is, however, a concept arbitrarily laid down by booksellers, book collectors and librarians. (The creation of the Federation of Central African States, which included the then Northern and Southern Rhodesia and Nyasaland,

led to certain material being classed as Rhodesiana, material which was later limited to Rhodesia – the Zimbabwe of today. Rhodesiana is regarded as a subsection of Africana.)

It may be said that Africana has to do with almost every aspect of South Africa: its history, climatic conditions, transport, architecture and the like. It is a very wide field of study, and includes records of every human activity. Only some of these records are important to librarians, while certain others are important to archivists or museologists. The first part of the discussion will deal with library Africana, including books, films, video recordings, sound recordings, ephemera and manuscripts.

The term 'Africana' is a creation of the twentieth century that came into fairly general use in 1920. Many authors divide Africana into different periods, with each one following his or her own system of subdivision. It is, however, possible to distinguish certain definite periods. For instance:

- The Portuguese period, or the pre-colonisation period: mainly accounts of journeys by Portuguese, Dutch and English travellers
- The Dutch period, 1652-1795: the *Daghregister* of Jan van Riebeeck and his successors (the best example of Africana of this period)
- The First British Occupation, 1795-1803
- The Batavian period, 1803-1806
- The Second British Occupation, 1806-1910
- The Great Trek and the founding of Natal and the Boer Republics of the Transvaal and the Orange Free State
- The Anglo-Boer War, 1899-1902
- The period of the Union of South Africa, 1910-1960
- The Republican period, 1961-1994
- The era of democracy, 1994-

The Great Trek and the Anglo-Boer War periods are often regarded as forming part of the First British Occupation. Moreover, many people regard Africana from 1910 onwards as being modern Africana, and only pre-1910 material is acknowledged as being true or classical Africana. The foregoing division is a purely chronological one. Africana can also be classified according to subject, for example records dealing with South Africa's fauna or flora, architecture, geography, history or accounts of journeys.

Types of Africana

As mentioned earlier, Africana can appear in any physical form. Most Africana is in the form of printed codices, as the discovery of the Cape of Good Hope (1487) came soon after the invention of printing in Europe (roughly 1450). There are 23 different kinds of Africana, based on the type of medium used for publication.

Books or Codex Records

Most Africana are in the form of printed books. In the early stages, these were printed in Europe, and some even in Batavia. They consist mostly of travelogues written by Portuguese,

English and French travellers, with a few in a Scandinavian language. They focus mainly on the unique plants, animals and people encountered during their travels. Carl Thunberg and Anders Sparrman (1747-1820), pupils of Charles Linnaeus (1707-1778), regarded as the father of modern botany, paid a visit to South Africa. When Linnaeus was drawing up his system for the classification of plants according to flower genus, he had access to a collection of Cape plants, very probably in Amsterdam where he worked and studied. His classification system was published in three works, *Systema naturae* (1735), *Genera plantarum* (1737) and *Species plantarum* (1753). Although Linnaeus himself never visited the Cape, he took a tremendous interest in Cape flora. Sparrman and Thunberg, however, are credited as the first to make an intensive study of Cape plants.

Before the art of printing was established in the Cape, codices were also written by hand. The Huguenot teacher Migault, for example, has bequeathed a hand-written Bible.

After printing became universal in the nineteenth century, locally printed books were produced. Many of these were in the form of government proclamations, missionary writings, sermons, law reports and several literary works. The First and Second Afrikaans Language Movements gave impetus to the writing of fiction. Early Afrikaans fiction is now much sought-after Africana. Books written in the vernacular form a valuable part of Africana. Beautiful illustrations are an important element of these codex records. The works of well-known artists such as Heinrich Claudius, François le Vaillant and, latterly, Cythna Letty's works, appear in codex form. (The codices have often been cut up, and the illustrations sold separately.) In many codices, excellent and reliable maps are found, such as those in *Travels in Southern Africa* by MKH Lichtenstein (1780-1857). An interesting characteristic of these works was that many of them were published in more than one language. This can be attributed to the great interest taken in the Cape because of the importance of its sea route.

Photographs and Films (Metacodex Records)

Photographs are highly rated as Africana material. They provide a picture of what South Africa looked like a hundred years ago. They have replaced common drawings in books as illustrative matter.

There were quite a few well-known photographers in South Africa. One of these was Arthur Elliot, who systematically took pictures of old buildings and unique events. His collection of 10 000 negatives was bought by the Union Government and is now housed in the Cape Archives.

Ephemera

Harrod (1977) defines the term as follows: 'pamphlets, cuttings and other material of ephemeral interest and value[;] [s]uch material of earlier periods that has acquired literary or historical value'.

This definition implies that the items concerned were of an ephemeral and fleeting nature at the time of their publication, but that they acquired value as Africana as time passed. Ephemera come in a variety of physical formats, and can fall under any of the

three previous subheadings. One finds printed ephemera such as pamphlets, handbills, bus tickets, Christmas cards and picture postcards. Interesting ephemera include theatre and circus programmes, and programmes of festive occasions.

Archival Records

Archival records are among our most important Africana records. Respective archival depositories in the ten provinces (Bloemfontein, Pretoria and Pietermaritzburg) and the National Archives in Pretoria (as well as Harare and Windhoek) contain most valuable sources. Kennedy (1965) is of the opinion that the entire history of the world since 1652 is mirrored in the Cape Town archives. There are many important collections of Africana in overseas archives, in places such as The Hague, Lisbon and London. In South Africa, specialised archives also exist such as political archives. The archives of private companies, such as De Beers or the Anglo-American Corporation, also contain valuable items.

Manuscripts

One important type of Africana is the handwritten manuscript. This includes letters, book manuscripts, diaries and the like. Many libraries such as the National Library of South Africa (Cape Town), universities such as Wits, Stellenbosch and Unisa, and museums such as the War Museum for the Boer Republics, have extensive collections.

Other

Africana could also include items such as furniture, vehicles, medals, coins and paper money.

Africana Collections

Before discussing the indexing of Africana, mention should be made of some prominent Africana collections. The librarians or curators in charge of these collections are usually knowledgeable about the field and about the indexing of Africana, and they could be contacted for more information through their specific institutions.

Many Africana collections were built up towards the end of the nineteenth and the beginning of the twentieth centuries. They consisted mainly of classical Africana and in many instances these collections formed the nuclei of the Africana collections found today in many of South Africa's larger libraries.

One of the best known of the Africana collectors was Sidney Mendelssohn (1861-1917). He once owned a diamond mine in Kimberley and later emigrated to England. He built up an extensive collection for which he also compiled a bibliography. On his death, he left his collection to the Union Parliament. Today it still forms part of the collection of the Library of Parliament.

There are several such collections in the National Library of South Africa in Cape Town, formerly the South African Library. Both the Dessinian and the Grey collections contain Africana. Another well-known collection is that of Charles Aken Fairbridge. His collection was transferred to the library in 1927.

Regarding Gauteng or the former Transvaal, one thinks mainly of John Gaspard Gubbins (1877-1935), Harold Strange, Sir EP Solomon and his son Hugh, and Sir Ernest Oppenheimer (1880-1957) and his son Harry, who all developed magnificent collections. The Gubbins Africana book collection is housed in the library of the University of the Witwatersrand, while other items from his collection form the nucleus of the Africana museum in Johannesburg. The Strange Collection is housed in the Johannesburg Public Library and the Solomon Collection in the Kosie Gericke (former Carnegie) Library at the University of Stellenbosch. The Oppenheimer Collection is still in the possession of the family where it is preserved in its own private library, the Brenthurst Library. Valuable unpublished material from this collection is published by the Brenthurst Press and these often set the standard for publishing and indexing of Africana in South Africa.

The Killie Campbell Collection is administered by the University of KwaZulu-Natal. This collection was built up by Margaret Roach Campbell (1881-1965). Some of the material in the library is now being digitalised as part of the Digital Imaging South Africa (DISA) project.

Other famous collections include the Dreyer Collection in the library of the University of the Free State and the JL van Schaik and Wagener Collections in the Unisa Library. Mention should also be made of the William Fehr Collection consisting of prints, watercolours and furniture. The porcelain and furniture are housed in the Castle in Cape Town, and the prints and watercolours in the historic homestead 'Rust en Vreugd' in Buitekant Street, Cape Town.

What Kinds of Africana are Indexed?

Africana material in various forms and formats may be indexed. For the purpose of this chapter, attention will be paid to books, journals and magazines.

Most libraries collect printed books, maps, art prints and metacodex records, whereas archives usually concentrate on collecting documents and other primary sources. Museums collect items such as ephemera, tools, furniture and vehicles.

Africana Indexing Resources

Indexes

There are many excellent indexes available for Africana material, for example the different indexes to *A South African bibliography to the year 1925, being a revision and continuation of Sidney Mendelssohn's South African bibliography 1910 (SABIB)* published by the State Library in 1985 (see detailed account in the second part of the discussion in this chapter) and Mansell in four volumes in 1979 respectively. *SABIB* differs from the South African National Bibliography (SANB) and the Retrospective South African National Bibliography (RSANB) in the sense that it also includes publications **about** South Africa. The fifth and sixth volumes were published in 1991 and 1997 respectively. The sixth volume is a comprehensive index to the whole bibliography and is an important indexing achievement in the South African context.

In 1983 the *Bibliography of anonymous and pseudonymous works from a South African bibliography to the year 1925* compiled by Dorothea Smit was published. A separate

volume listing South African printers and publishers from 1795-1925 was also published in 1987 by the South African Library (*South African printers and publishers 1795-1925)*. The oldest item printed in South Africa dates from 1795. This publication makes SABIB even more accessible through a number of important indexes.

SABIB is an update of Mendelssohn's bibliography *South African bibliography* in two volumes published in 1910, but without the numerous annotations in it. Indexers still find it a useful reference source.

The indexes to SABIB are excellent examples of how an Africana index should be created and should be consulted by prospective and practising Africana indexers.

Reference Sources

The Africana indexer in South Africa is in the fortunate position that there are a large number of reference sources on all aspects of South African Africana, from history to science, to the fauna and flora of South Africa and to its people:

- One of the best is the *South African biographical dictionary / Suid-Afrikaanse biografiese woordeboek* (1987). It was updated by a further two volumes which, unfortunately, are not of the same quality as the first five. It has an excellent name index.
- Mention should be made of indexes to streets in or general histories of Southern African cities. Examples are Tom Andrew's *Street names of Pretoria* and *A concise historical dictionary of Greater Johannesburg* by Naomi and Reuben Musiker. A useful index to South African place names was compiled by PE Raper in 1987, entitled *A dictionary of Southern African place names.*

Problems with Indexing Africana

The indexing process of Africana follows the same steps as any other indexing procedure; namely description, conceptual analysis, assigning indexing terms, et cetera. These do not bear repetition here but some problems peculiar to Africana indexing will be discussed.

Handwriting

The earliest Africana is usually handwritten and written in seventeenth century Dutch. These include diaries, travelogues, correspondence, and so on. The indexer should be able to read them and transcribe them into modern Dutch or into the language of the index being compiled.

Language

Cape Town, as the 'Tavern of the Seas', was visited by many Europeans who often wrote an account of their visits – such as Barrow, Liechtenstein and Le Vaillant. It is therefore imperative that Africana indexers should have a sound command of several languages, such as Portuguese, French, Dutch and German. Knowledge of the indigenous languages of South Africa is another prerequisite for the Africana indexer.

Historical Names of Persons

In true colonial tradition the names of the indigenous peoples of Africa were Europeanised. One of the first local citizens to be met by Jan van Riebeeck in 1652 was Autshumao or Hadah, captain of the Khoi tribe the Goringhaikona, who lived in the vicinity of Table Bay. Autshumao met English travellers who renamed him Harry or King Harry. For hundreds of years this is how he was known to all.

What name should the indexer use if the document uses only Harry or Hadah, or both? The best modern practice would be to use the indigenous name and to refer by *see* references from Harry to Autshumao. There are thousands of such examples in Africana books. SJP Kruger, the last President of the South African Republic, was known as 'Oom Paul', simply as Paul Kruger, or as 'die ou President', et cetera. Which term should be used in the index? One could opt for Kruger, Stephanus Paulus Johannes, President of the South African Republic, 1884-1902, with the necessary cross-references.

Place Names

Place names create the same kind of problems for the indexer of Africana. In South Africa, place names may have an Afrikaans, English and/or vernacular name. The Witwatersrand is known by this name in Afrikaans and English, the Zulu name is Egoli and the Sotho name Gauteng. Previously, Gauteng was called the PWV area (Pretoria, Witwatersrand Vereeniging area) and before that known as Southern Transvaal. The new province established after 1994 could be indexed under Gauteng with references perhaps from PWV, Egoli, Southern Transvaal and a general reference from Transvaal *see also* Gauteng, Limpopo, Mpumalanga and North West province. A book dealing with the gold mines of Johannesburg in 1894 should be indexed under Johannesburg, with references from Gauteng, Egoli, PWV and Witwatersrand.

Every book or periodical article should be treated on its own merit. During conceptual analysis, the indexer should determine the main subject of the book, as well as additional subjects that warrant indexing terms.

Different Names for Different Events

The problems are the same as above. The Anglo-Boer War is known as the Boer War, the South African War, 'die Engelse Oorlog', 'die Tweede Vryheidsoorlog', Second War of Independence and, at the time of the war's centenary, the Anglo South African Boer War. In such instances (as with personal and place names, mentioned above) the best solution is to create a standardised form of name, linked through cross-references to the other name forms.

Changes in Orthography

Orthography refers to the spelling of words. A number of years ago, the orthography of Sesotho changed. Recently Afrikaans also underwent a radical change where words that were once written as one word can now be separated into two, for example Tafelberg, the

old spelling, may now be written Tafel Berg. A living language changes constantly, and British English is widely used in South Africa, not the American version.

Language Used in an Index

In Africana indexing, the language of the index can be a problem. As a general guideline the index should be compiled in the language of the book. A library catalogue can be problematic. Some libraries use English subject headings only; some use both Afrikaans and English. One should be guided by the purpose and users of the index in context with the indexing policy.

South Africanisms

In South African English and Afrikaans, certain typical South Africanisms are used, for instance 'robot' instead of 'traffic light'. Other examples are 'imbizo' or 'bosberaad' for the American 'brainstorming', 'makietie' for a party, and so on. South African terms should be used in the index if they appear in the text, but references should be made from the more acceptable terms.

Weights, Measures and Monetary Systems

In older Africana, information sources such as pounds and ounces are still used, as well as miles, yards, feet and inches. Even older books have references to the Dutch monetary system of riksdaalders (rixdollars), roede, vaam, et cetera. The English pounds, shillings and pence were used until 1961 when South Africa changed to the decimal system of rands and cents.

Use the old measurements for old documents but refer to the new ones as well. A good idea is to include a calculation table, for instance 2 pounds and 2 ounces = 1 kilogram, or 1 shilling = 10 cents.

A Good User Guide

The problems mentioned above illustrate the need for a thorough introduction and user guide to any Africana index. Terms or measurements used should have an explanatory note, with calculation tables provided and, where necessary, annotations. Important to Africana indexing, a controlled vocabulary or thesaurus should be used and policy decisions should be taken regularly.

Conclusion

Africana is an important information source and it is therefore essential that Southern African indexers should have a sound knowledge of the subject, its indexing tools, reference works and methodology.

The second part of this chapter is devoted to the South African Bibliography (SABIB).

SABIB: A Jewel in the Crown of South African Bibliography and Indexing

In WS Gilbert's *Mikado*, Ko-Ko, referring to his enemies, says,

'I've got a little list' – to which the chorus replies:
'He's got 'em on the list – he's got 'em on the list;
And they'll none of 'em be missed!'

Historical Background

This parody of a concept which librarians have based their work on for centuries nonetheless points to the various uses to which the list can be put. In library terms it is the foundation of the work of the cataloguer, bibliographer and indexer. A definition of indexing is '[t]he compiling of systematic guides to the location of words, names and concepts in books and other publications' (*Cambridge encyclopaedia* 1994). The Latin word 'index' in its dictionary definition of 1598 is referred to as a pointer, a discoverer, which serves to direct to a portion, fact or conclusion. The concept of the index is therefore an old one that goes back to classical times. Cicero used the term to describe the list of contents to a book. Claudius Ptolemy's great astronomical and trigonometrical work, the *Almagest*, in the second century AD, was a model of organised tables and lists of stars and planets, unsurpassed until the fifteenth century. The cartographers, astronomers, mathematicians and geographers of the Renaissance were much given to producing lists of numbers, place names in gazetteers, and the logic of alphabetical order was accepted. In *Troilus and Cressida*, Shakespeare wrote:

And in such indexes, although small pricks
To their subsequent volumes, there is seen
The baby figure of the giant mass.

Another example of use of an index was direct and to the point. The *Index librorum prohibitorum* (1559) was inaugurated and published by Pope Paul IV, formerly Cardinal Carafa, the head of the Papal Inquisition. The *Index* listed books which were prohibited reading matter for Catholics. The origin of this kind of restriction dates back to 496 AD. The Catholic Church continued to publish new editions of the *Index* until 1966, when it was decided that although the papacy still claimed the right to restrict the reading available to Catholics, no further lists would be published.

If the index in its widest frame of reference is seen as a list of items, the lines dividing a bibliography, a catalogue and an index are somewhat blurred. At times they seem to fulfil the same function – as devices to find things which have been arranged in alphabetical or subject order. In the field of Africana this is amply illustrated by that indispensable tool for the librarian and researcher, the six-volume *South African bibliography to the year 1925: being a revision and continuation of Sidney Mendelssohn's South African bibliography (1910)*. The six volumes combine all three of the elements mentioned above.

The background to the compilation of the bibliography, known as SABIB, is interesting. In 1882, the historian George McCall Theal suggested that there was a need for a bibliography of South African historical material. He himself, much later in 1912, produced his annotated

463-page *Catalogue of books and pamphlets relating to Africa, South of the Zambezi* in Cape Town. By this time, two other bibliographies had appeared. The first, by CA Fairbridge and John Noble, entitled *Catalogue of books relating to South Africa* 1886, was based on Fairbridge's own collection. The second was HC Schunke-Holloway's 'Bibliography of books, pamphlets, maps, magazine articles et cetera: relating to South Africa, with special reference to geography from 1503-1888' (*Transactions of the South African Philosophical Society* 1898). Although these were admirable attempts at forming early records of Africana, they were really lists with no annotations and scant bibliographic information.

Subsequent Developments

The advent of Sidney Mendelssohn's *South African bibliography* in 1910 provided a much-needed resource in the field of descriptive cataloguing of Africana. The massive two-volume work, a bibliography of Sidney Mendelssohn's own fabulous collection, contained 7 000 titles, many of them accompanied by useful annotations. For the next 20 years Mendelssohn was the bible for librarians, researchers and collectors. The collection was later bequeathed by Mendelssohn to the Library of Parliament. By 1934, the time had come for a revision to cater for the needs of the growing number of libraries and other interested users. World War II interrupted the initial plans and it was not until 1957 that a facsimile reprint of Mendelssohn was issued. All 500 copies were quickly sold. Its value was by then well established in the book world. Not in Mendelssohn, in a library situation or in a dealer's catalogue, usually indicated a rare item of Africana dating from earlier than 1910. The facsimile proved to be only a temporary solution to the problem of putting together the mass of material which had accumulated over the years.

In 1960, a serious revision was begun by the South African Library in Cape Town (then the South African Public Library). In order to make the scope of the vast amount of material manageable, it was decided to exclude pictures, periodicals, newspapers, maps, printed music, manuscripts and government publications after 1854. The geographical scope would be confined to the area south of the Limpopo River, but would include the three British High Commission territories and those islands of the Indian and Atlantic oceans that were once British colonial possessions. On the face of it, this would appear to eliminate a large body of material, but by 1960 the number of books and pamphlets concerning South Africa from 1501 to 1925 comprised enough material for a very large bibliography. Presumably the hope was that the categories excluded would be separately dealt with by specialists. Much has been done to cover this ground, but there remains a great deal of material to be worked on.

The format of SABIB took the Mendelssohn Collection as its basis, with the entries arranged by author. To these were added the catalogues of 29 libraries and private collections. Each entry provides the source or sources, depending on rarity, of the item, making the initial four-volume bibliography an invaluable finding list. A supplement was added in 1991, and by 1997 a subject and title index appeared as the sixth volume, effectively rounding off the project. The cataloguing was meticulously and imaginatively done throughout, making the bibliography, in today's terminology, user-friendly. The index was compiled using the Macrex computer program which successfully dealt with the 15 000 main items in the subject section, with subheadings and cross-references. The title index has some 35 103

titles, also with some cross-references. The subject contents of the staggering number of 40 000 books and pamphlets are contained in the index. Some of the subject divisions have an enormous number of entries, for example for Portugal across the five volumes. To break a section like this into sub-headings would have made the work tremendously difficult for the compilers. Even so, the usefulness of the index far outweighs the effort of working through a large subject section.

As a combination of catalogue, bibliography and index, SABIB fulfils many functions, depending on the requirements of the user. From the point of view of the librarian, bibliographer and indexer, to be able to compare editions, establish dates and pinpoint details like number of plates, and many other vital elements dear to the heart of the true bibliographer, is a boon indeed.

Other major tools in South African bibliography and indexing, some of long standing, and which are now being produced electronically, are briefly referred to below.

The Fairbridge Index

The Fairbridge Index, compiled by the Rhodesian pioneer journalist WF Fairbridge (1863-1943), is a manual one written on cards, in the South African National Library (Cape Town). Unfortunately Fairbridge's handwriting is not very clear. The JG Gubbins Africana Library at Wits University and TAB (National Archives Repository: Records of the Old Transvaal) have photocopied versions of it. This very random index is based on clippings from the *Cape Argus,* other Cape journals and some books. It covers Africana from around 1800 to the beginning of the twentieth century. Subjects include geography, 'natives' (a term of its time, referring to blacks) and shipping. It contains useful references to places and people. Staff who did the photocopying for Wits Library estimated that it extended to 60 000 traditional catalogue cards (Strachan 1985: ms). This valuable resource needs to be enhanced to an electronic format.

Index to South African Periodicals (ISAP: 1940-1986)

ISAP follows Fairbridge. It covers the period 1940-1986 and came out in various formats, printed and microfiche. Most South African journals were included. It started as a group effort under the aegis of the South African Library Association; very soon Johannesburg Public Library took over compilation and publication. Since 1987 to date, the National Library of South Africa is producing the online electronic version. This version, obtainable through the National Library's web page www.nlsa.ac.za, as well as SABINET Online www.sabinet.co.za, usually seems six months behind the current date. Further details about ISAP can be found in chapter 9.

South African National Bibliography (SANB: 1959-1999)

From 1959, SANB was published annually until 1999, covering only material published in South Africa. It has been continued as an online publication still known as SANB and also obtainable through the National Library's web page, using its own database

and SABINET Online. Following on from SABIB, with its coverage until 1925, the period 1926-1958 is covered by the two-volume *Retrospective South African national bibliography* (RSANB:1985). The various lists that had covered this period were considered unsatisfactory, so RSANB was published in 1985.

African Studies and *South African Studies*

These two electronic sites were begun when ISAP ceased to be published in printed format. They can be traced through the parent web page www.nisc.com and may be accessed in both CD-ROM and online (Biblioline). *African studies* claims cover from the nineteenth century to the present and *South African studies* from 1960 to date, with some nineteenth century cover. A trial run for the nineteenth century found them to be wanting in some respects. For authority, their subject index could be useful. Their best usage is for the post-1987 period to date. These sites are not limited to material published in South Africa.

South African Freedom Struggles, 1950-1994

DISA publish this full-text database. Its aim is to make 'Southern African material of high socio-political interest, which would otherwise be difficult to locate and use, [and] accessible'(http://disa.nu.ac.za).

Forty journal titles are covered in the database, dating 1950-1994. Many viewpoints, both black and white, are represented; from the *African Communist* to the journal of the Liberal Party of South Africa's *Contact*. Many of the titles are not well represented in research collections. The database does make use of a thesaurus.

Authority for Indexing and Thesauri

All of the printed and electronic databases mentioned above, except maybe the Fairbridge Index, can be used for these purposes. The National Library of South Africa administers an accurate system for South African corporate and personal names. The Authority File of SACat, found through SABINET, may be used for subject searches.

BIBLIOGRAPHY

Americana. 1981. *The Book Collector*, 30(4):447-460.

Andrews, TE. 1989. *Street and place names of old Pretoria.* Pretoria: Van Schaik.

Bibliophilia Africana. 1967. Cape Town: Friends of the South African Library.

Bibliophilia Africana. 1978. Johannesburg: Friends of the University of the Witwatersrand Library.

Bibliophilia Africana IV. 1981. Cape Town: Friends of the South African Library.

Boekspieël van Suid-Afrika. 1952. Kaapstad: Komitee vir Boekuitstalling Van Riebeeckfees.

Bradlow, FR. 1970a. Africana. *Standard encyclopaedia of Southern Africa*, Vol 1. Cape Town: Nasou: 46-57.

Bradlow, FR. 1970b. *Africana books and pictures: a selection of published papers.* Cape Town: Balkema.

Cambridge encyclopedia. 1994. 2nd ed. Cambridge: Cambridge University Press.

Colvin, ID. 1979. *Introduction to Africana: being a newly illustrated reprint of the Introduction to the South African bibliography by Sidney Mendelssohn.* Cape Town: Balkema, for the Friends of the South African Library.

Fairbridge, WF. [19–]. Fairbridge Index. Unpublished ms. Available: http://disa.nu.ac.za

Funk & Wagnalls standard dictionary of the English language. 1968. International ed. New York: Funk & Wagnalls. 2 vols.

Godfrey, D. 1963. *The enchanted door.* Cape Town: Timmins.

Gordon-Brown, A. 1975. *Pictorial Africana.* Cape Town: Balkema.

Gordon-Brown, A. 1979. *The Settlers' Press.* Cape Town: Balkema.

Harrod, LM. 1977. *The librarians' glossary.* 4th rev. ed. London: Deutsch.

HAT: verklarende woordeboek van die Afrikaanse taal. 1983. Redakteurs: FF Odendaal ... et al. Johannesburg: Perskor-Uitgewery.

Kennedy, RF. 1965. *Africana repository.* Cape Town: Juta.

Musiker, R. 1980. *South African bibliography: a survey of bibliographies and bibliographical work.* 2nd ed. Cape Town: D Philip.

Musiker, N & Musiker R. 2000. *A concise historical dictionary of Greater Johannesburg.* Cape Town: Francolin.

Pama, C. 1968. *Die Suid-Afrikaanse Biblioteek: sy geskiedenis, versamelings en bibliotekarisse, 1818-1968.* Cape Town: AA Balkema.

Pretoria Africana Association. 1975. *Yearbook 1,* 1975. Pretoria: The Association.

Pretoria Africana Association. 1977. *Yearbook 2,* 1976. Pretoria: The Association.

State Library. 1985. *RSANB, 1926-1958.* Pretoria: State Library.

Raper, PE. 1987. *A dictionary of Southern African place names.* Pretoria: HSRC.

SANB: South African national bibliography. 1959-1999. Pretoria: State Library.

Schunke-Holloway, HC. 1898. Bibliography of books, pamphlets, maps, magazine articles et cetera: relating to South Africa, with special reference to geography from 1503-1888. *Transactions of the South African Philosophical Society,* 1(2):131-294.

Smith, AH. 1973. *Africana curiosities.* Johannesburg: Donker.

Smith AH. 1976. *Africana byways.* Johannesburg: Donker.

South African bibliography to the year 1925: being a revision and continuation of Sidney Mendelssohn's South African bibliography (1910). 1979. Edited at the South African Library, Cape Town. London: Mansell (4 vols. Supplement (vol 5). Indexes (vol 6).

South African biographical dictionary / Suid-Afrikaanse biografiese woordeboek. 1987. Pretoria: HSRC.

A world too wide: essays on libraries and other themes in honour of Reuben Musiker. 1993. Johannesburg: University of the Witwatersrand.

18 ANTIQUE MAPS

Graphic Delights: Landscapes Unfolded on Antique Maps

Yvonne Garson

Abstract

The aim of the author is to give an impression of the responsibilities and difficulties, as well as the aesthetic enjoyment, inherent in the custodianship of a collection of antique maps. Also assessed is the educational benefit to the librarian undertaking this task. Some practical points are made concerning the care, cataloguing and indexing of old maps, in sheet form as well as in books, bearing in mind the needs of the institution concerned. The reference work required is also discussed. Some historical background is supplied, together with a list of suggested reading.

Some Thoughts on the Management of Collections of Old Maps

As this chapter's title suggests, there is much in the study of the antique map to stimulate the imagination and please the eye. The broad scope of the subject encompasses the earliest attempts, thousands of years ago, by human beings to make simple graphics to establish their location, the new cartography of the Renaissance, the glorious elaborations of the Dutch and Flemish maps of the seventeenth century, and the more accurate and elegant maps of the eighteenth and early nineteenth centuries.

Within the many fields of specialisation in the discipline of librarianship, which today seems to be weighted heavily towards a technological approach, there is fortunately still room for the librarian with an antiquarian cast of mind. Collections of all kinds of old objects, books, manuscripts, maps and other graphic material are still being unearthed in odd corners of the world, which require specialised treatment. New technology is an invaluable tool in the management of old material, but it cannot take the place of understanding and evaluating it to its best advantage. This chapter is presented against the background of the Africana map collection in the Library of the University of the Witwatersrand, Johannesburg. The intention, however, is to focus on broader insights gained from the bibliographic and indexing experience in the recording of this collection, in the hope that some useful guidelines can be provided for the management of antique maps.

By the standards of the great institutional collections such as those in the British Library, the Library of Congress in Washington and the Bibliothèque Nationale in Paris, the Africana collection of maps at Wits is a small one. But through examining its contents it is possible to trace the progress of cartography from the fifteenth to the nineteenth centuries, supported by a variety of reference books including facsimile atlases of the works of Claudius Ptolemy

and others of the great scholars of the history of mapping. Over the years additions have been carefully made to enhance the quality of the collection. On setting out to build up and expand a good collection of historical maps, quality above quantity should be the guiding factor, naturally within budgetary limits.

The custodianship of a collection of old and valuable material brings with it heavy responsibilities. There is the physical care of often brittle and fragile items which need special treatment. There is also the need to secure one's material against theft. A few years ago the Wits Library suffered serious losses through both burglary and ingenious pilfering. Some of the maps were recovered, but regrettable gaps remain. Watertight security is of paramount importance, but even the most sophisticated systems have been known to have been breached from without and within. In 2003 Scotland Yard estimated that approximately 4 500 maps were missing from European Libraries alone. An article in the *Times Literary Supplement* entitled 'Don't steal this book' gives these alarming figures and also some illuminating facts about the history of this specialised type of plunder over the centuries (Caines 2003).

The system of care of precious items such as old maps has of necessity to be adapted to local conditions. In an ideal world, climate control, special lighting and perfect storage space would all be possible. Few libraries today are able to provide an ideal environment, but certain basic measures should be taken if possible. These can be summarised as follows:

- Encapsulate all sheet maps in acid-free plastic sleeves. This protects them and facilitates handling.
- Store in flat drawers in cabinets that must be protected against fish moths and damp. Some libraries have special cabinets in which the maps are hung on rods, the upper orders of the maps being reinforced. In some ways this is a good method since it is easier to keep the maps in an ordered system, but it is probably better to allow them to lie flat.
- Keep lighting as low as possible. This applies as well to framed maps, which should always be backed and mounted with acid-free boards.
- Keep the work of restoring old maps to a minimum. Tears can be repaired with special museum-approved tape. Beyond this they are better left alone, bearing in mind that heavy restoration greatly reduces the value of any old material, and in some cases can prove to be damaging to the images.

Once the physical environment of a map collection has been satisfactorily established, the next stage is the evaluation of the items for recording.

Evaluation of Maps for Recording

Of all the antique graphic materials which come the way of the librarian, maps are among the most interesting. One is confronted here not only with form, which in many cases is beautiful in outline, decoration and lettering, but with content as well. Content can often prove to be baffling; maps are frequently enigmatic in the amount of information they yield. What needs to be understood is that even if the outline, as in many a map of

the sixteenth century, for example, seems to be wildly inaccurate, the finished product nevertheless represents a synthesis at a given time of the work of philosophers, historians, mathematicians, astronomers, geographers, mariners, travellers, explorers, engravers and illustrators. The aesthetic quality can also range widely from crude and harshly coloured items to exquisite works of art, delicately engraved, lettered and coloured. French maps of the eighteenth century stand out in this last category. Maps of the old kind, as opposed to the impeccably accurate maps of today, are often graphic story-tellers of more lurid times, particularly when accompanied by text, as for example in some of the sixteenth to eighteenth century atlases. From these come images of piracy, slavery and colonial inroads into conquered territory. The antique map is not always just a pretty face.

The history of cartography stretches back through the centuries, and to do one's collection justice, the rudiments at least should be absorbed, and an ability developed to read the features of the antique map for clues. There can be few librarians, however erudite, who are qualified in astronomy or navigation, for example, but there are many excellent histories which offer accessible background information. The study of aspects of cartographic history is ongoing; the fascination with the antique map and atlas never ceases.

Structural Cataloguing

The structural cataloguing of a map can be a reasonably simple exercise. The rules suggested by the second edition of the *Anglo American cataloguing rules* (1988) provide a useful framework to follow. Antique maps are by their nature quixotic, and one must be pragmatic in the application of rules, with an eye to consistency if adopting a style to suit a particular collection. The general order of initial cataloguing runs as follows: author followed by dates, title, scale, edition, imprint, size in centimetres and an indication whether the map is coloured or not.

Provision is made for notes at the end of the entry. An example would be

Senex, John d.1740
Africa corrected from the observations of the Royal Society of London and Paris ...
[London], John Maxwell, [1711-1714]
Scales vary
1 map:col. 69 x 97 cm.
Notes

Depending on the nature of the library or museum, arrangement of the material can be either by author or area or date. Where the historical element rather than the geographical predominates, the most practical way to arrange the catalogue seems to be chronologically, with authors alphabetically disposed within the chronology. Added entries for additional elements such as engravers, illustrators, publishers, if they are known to be of importance, and any other salient features, should be made.

Certain complications sometimes arise in the cataloguing of historical maps. Their enigmatic nature has been referred to. On the majority of old maps, the cartouche, often a work of art in itself and frequently highly symbolic, contains at least some of the essential

bibliographic information. It can happen though that there is neither author, imprint nor even a title. In such a situation research has to be undertaken which can become very time-consuming. Reference material is invaluable here and there is a wide field of sources to consult. Experience has shown that one of the most helpful of these has been the British Museum's *Catalogue of printed maps, charts and plans.* There is also the Library of Congress *National Union catalogue*, and of course exploration on the Internet.

The starting point is to try to establish the identity of the author, utilising all available clues on the map itself, and employing the use of illustrated reference material from which comparisons may be made. This can often open up the whole enquiry, and give enough clues for speculative names, dates and places of publication. In the extreme case of a totally anonymous map, the watermark on the paper can at least pinpoint the item to within a certain date range. Colouring can also be followed up, but with caution. This is one of the most difficult areas when it comes to dating. Modern methods of antiquing in paper making and colour reproduction can cause one to make false claims, and the advice of experts should be called in. It is necessary to be watchful for telltale signs of copying. These are usually indicated by faint platemarks and colour, obvious erasures made here and there and the superimposing of new information in the cartouche. The usual explanation for this type of tampering is that an unscrupulous publisher, having acquired or even stolen plates of a famous map or maps, has reprinted the map, usually without change to the content, sometimes as long as a hundred years after its original publication. Here is another intriguing avenue of investigation which the fascinating and recalcitrant old map can present. As is the case in general cataloguing practice, externally acquired information must be placed within square brackets. Once the bibliographic details have been put in place, the notes can be assembled.

The scope of the note depends on many factors. When dealing with old maps it is important to situate them within their historical context. If the cartographer is known, biographical details as well as his place in the development of cartography should be included. Information about famous cartographers associated with the development of mapping in Europe from the time of the Renaissance, such as Mercator, Ortelius, Sanuto, Sanson, de l'Isle, d'Anville and a host of others, is readily available, should one be so fortunate as to have their maps in a collection. It is with the more local maps of an area that other methods need to be explored to find details of their makers and the intent behind their construction. In the Africana context, with the opening up of regions and districts by colonial administrators, explorers and missionaries, one encounters maps of greater detail and accuracy than the grand and lavish Dutch and French maps of the seventeenth and eighteenth centuries. Because these have bearing on local matters, they require careful analysis and description. They are often the subjects of research in various fields and need to be accessible. This is where good notes are important. If at all possible a printed catalogue, even if only for the use of the institution, should be compiled. Computer records, mandatory today, are necessarily limited in scope – which can be frustrating to the cataloguer who may feel it necessary to provide extensive information. The ideal solution to the thorough recording of important maps, whatever their emphasis, is to compile a printed catalogue, with an index, even if it is only for the use of the particular institution. To provide such a

tool would greatly benefit the librarian and the researcher. Experience has shown that the most unlikely-seeming details can be of interest and value.

Publication

If publication is thought to be desirable, the system of ordering the entries as described throughout the chapter has proved to be practical. In the case of publication of the Wits Africana map collection, of which three bibliographies have been completed, all the inherent problems with maps, be they manuscript or printed, were encountered; this type of compilation is very time-consuming. In addition, because the index included personal and place names in old Dutch, German, Italian and French, it was not possible to adapt any existing indexing computer program. The index had to be put together item by item, a laborious procedure, but producing a satisfactory outcome. Illustrations to the text as well as an accompanying CD-ROM can enhance the usefulness of the publication, although the illustrating of a map catalogue is in itself an undertaking fraught with problems. Some maps are very large, the paper is often yellowed and brittle, and lines are often faint, making reproduction difficult.

An additional system that was found desirable for rounding out the collection of maps in the Wits Africana Library, which also contains fine Africana books, was to analyse some of the rarer items illustrated with maps by well-known cartographers, and to incorporate the maps in the bibliography. There are guidelines on how to set out this type of entry. Because of the introduction of additional bibliographical detail into the body of the physical description of these items, and also more extensive notes, the problem of compression of the added information was not easy to overcome; neither was it entirely successful in all cases. Notes can easily run away with one and the succinct paragraph or two is an ideal to aim for.

Conclusion

For the librarian who is in the fortunate position of recording and caring for a collection of historical maps for the first time, a short list of useful publications is included at the end of this chapter. It is to be hoped that the work and background reading required, as well as the handling of the material, will stimulate an appreciation of the aesthetic nature and historical interest of the antique map, at the same time providing the educational benefits of exploring a new field of librarianship.

Bibliography

Anglo-American cataloguing rules. 1988. 2nd ed, 1988 rev. Ottawa: Canadian Library Association.
Bagrow, L. 1964. *History of cartography.* Edited by RA Skelton. London: Watts.
Brown, LA. 1951. *The story of maps.* London: Constable.
Caines, M. 2003. Don't steal this book. *Times Literary Supplement,* August:11.
Crone, GR. 1962. *Maps and their makers.* London: Hutchinson University Library.
Fordham, G. 1969. *Studies in carto-bibliography.* London: Dawson of Pall Mall.

Garson, Y (comp). 1998. *Africa in Europe's eyes: maps of Africa in the Library of the University of the Witwatersrand, Johannesburg.* Johannesburg: University of the Witwatersrand.

Garson, Y (comp). 2003. *From myth to reality in the cartography of the colonial era: historical maps of Southern Africa and the islands off its coasts in the John G Gubbins Africana Library, University of the Witwatersrand, Johannesburg.* Johannesburg: University of the Witwatersrand Library.

Harvey, M. 2001. *The island of lost maps: a true story of cartographic crime.* London: Weidenfeld & Nicolson.

Imago Mundi: *The Journal of the International Society for the History of Cartography.* Published annually.

Koeman, C. 1964. *The history of Abraham Ortelius and his Theatrum Orbis Terrarum.* Lausanne: Sequoia.

Koeman, C. 1970. *Joan Blaeu and his Grand Atlas.* London: Philip.

Lynam, E. 1953. *The mapmaker's art.* London: Batchworth.

Moreland, C & Bannister, D. 1983. *Antique maps.* 3rd ed. London: Phaidon.

Norwich, I. 1997. *Norwich's maps of Africa: an illustrated and annotated carto-bibliography.* 2nd ed. revised and edited by JC Stone. Norwich, Vermont: Terra Nova.

Parry, JH. 1964. *The age of reconnaissance.* New York: The New American Library.

Shirley, R. 1983. *The mapping of the world: early printed maps, 1472-1700.* London: Holland.

Skelton, RA (ed). 1964. *History of cartography.* London: Watts.

Sobel, D. 1995. *Longitude.* London: Fourth Estate.

Tooley, RV. 1969. *Collectors' guide to the maps of Africa and Southern Africa.* London: Carta.

Tooley, RV (ed). 1979. *Dictionary of mapmakers.* Tring: Map Collector Publications.

Woodward, D. (ed). 1975. *Five centuries of map printing.* Chicago: University of Chicago Press.

19 ARCHIVES

Michele Pickover, Carol Archibald
and Kate Abbott

Abstract

In the not-so-distant past it was commonplace for researchers and users of archives to rely primarily on the knowledge and sleuthing abilities of individual archivists, if they wanted to gain access to archival records and their content. This was because archivists tend to stress documenting context and structure from the point of view of provenance, series and record-keeping systems. However, nowadays archivists are also seriously considering questions relating to the analysis and retrieval of information in archival records, particularly in relation to archival description that identifies, manages, locates and interprets archival holdings and clarifies their context and origin. Users' needs, the nature of the records and the principles underlying indexing itself are now taken into account in the setting up of a system.

Introduction

In the archival environment, the depth and structure of information is complex and multi-levelled. It is crucial, when analysing archival records for indexing purposes, to retrieve information in a way that still reflects their provenance and methods of organisation. If a well-planned finding aids system is not developed for every archival collection (which contains sub-collections and items to which additional material may be constantly added), and it is not comprehensively and carefully indexed, information will be lost, accessibility will be compromised, researchers will be disadvantaged and societal memory will be depleted.

The process of cataloguing archival material is very different from cataloguing books. An archives does not have an all-embracing subject-based catalogue to its holdings. To navigate the archival records researchers are dependent on finding aids. Finding aids, which include guides, inventories and indexes, provide for physical and intellectual control and make it possible to identify and retrieve specific record items, thereby facilitating access. Finding aids also assist in the preservation of archives by reducing the amount of physical handling of documents.

Indexes are the components of the descriptive archival system that allow researchers to link their searches to the hierarchically structured finding aids. However, because of the complexity and layered nature of archival description types, indexes are important tools to link more deeply to content. Entries reflect the names of record creators, series, items, personal and institutional names and subjects. It is therefore important to produce in-depth finding aids. Indexes can be in the form of cards, printed lists, databases and in-house and

web-based automated retrieval systems and should also allude to the hierarchical structures within the finding aids. They should also refer researchers to the context of an item, thereby increasing understanding or pointing to additional material. Consequently, indexing archival records can be labour-intensive, expensive, variable and subjective.

Each archive tends to have its own in-house rules with regard to indexing conventions and terminology used. However, the way indexing terms are selected and standardised is very important in an archive and good quality index terms are the best way of providing multiple access routes to information. In the South African historical and political framework, a thesaurus of preferred terms should be used and continuously updated so as to make indexing manageable and effective and to provide for consistent retrieval. Descriptions of archival collections are usually written at the time the material itself was created and therefore words and phrases used often reflect common usage of the era in which they were written. In the early 1990s, a group of archivists and librarians, largely from the nongovernmental organisations (NGO) sector, came together as the South African Thesaurus Building Group (SATBG) to develop a thesaurus that would be useful in resource centres and archives whose holdings were of a political, social and historical nature. Subsequently, Digital Imaging South Africa (DISA), a national collaborative digital project, has been systematically adding to the work of the SATBG.

In an archives, index terms are also key in accessing, searching and retrieving digital objects. A list of indexing protocols is developed to assure consistency in relation to the indexing terms and a system of review and quality control for all indexing tasks is used.

Painstakingly preserving records for historical and conservation reasons is fundamental, but if researchers over time cannot gain access to these documents or their contents, their value and power is lost. The production of finding aids and indexes to these records is indispensable to unlocking the potential of and gaining access to archival material.

The benefits to archivists and users derived from electronic and digital technology in relation to the production of archival finding aids are huge. It is more functional, allows access to the full details of collections and also enables those details to be shared with remote users though a World Wide Web (WWW) interface. An electronic finding aid, therefore, facilitates vigorous searching, retrieval and displaying of information because it can index across collections and provide full-text and item-level indexing within collections.

To ensure an effective global retrieval system and to reflect the complexity of archival description, while taking into consideration context, content and structure, the archival profession is currently developing internationally recognised descriptive metadata schema, such as Encoded Archive Description (EAD) and the Text Encoded Initiative (TEI). These are descriptive mark-up standards based on Standard Generalized Mark-up Language (SGML), which is a very useful tool for cataloguing finding aids. It is non-proprietary, flexible and can handle hierarchical structures. SGML allows refined and intelligent database and document indexing and searching.

Requirements for the encoding standard include the following criteria:

- ability to represent finding aids
- preservation of hierarchical relationships
- ability to inherit description from one level of the hierarchy to another

- ability to move among different hierarchical levels
- support for element-specific indexing and retrieval

Indexing Protocols in Historical Papers, University of the Witwatersrand

The University of the Witwatersrand has, as one of its most valuable and prestigious heritage and research assets, the holdings of the priceless Historical Papers collections. Historical Papers is the main humanities archival research resource on campus and is located in the William Cullen Library. It is also the largest non-state archives in Southern Africa.

Archives provide the bedrock for society's understanding of the past. They underpin citizens' rights, assert identities and are crucial to truth recovery. They are also irreplaceable evidential testaments of human experience on which social equality is built. The apartheid past, in particular, has left its mark not only in human suffering but also in the great volume of documents that record both the violations of human rights and the struggles to defend them. Many of the most valuable primary sources relating to human rights struggles are often in danger of being lost, either as a result of the dispersal of documents or lack of capacity of the human rights NGOs that hold them.

With this in mind, Historical Papers was established in the 1960s as a centre where South African memory, particularly that which was/is under threat or may be lost, is preserved and made available. We have concentrated on documenting and providing enduring evidence of resistance struggles against oppression and exploitation. As such, the voices of those who have been politically and socially marginalised are also reflected. The archives held in custody for the wider community within Historical Papers are vast and provide a unique, and often fragile, documentary record of South African history and culture, particularly with regard to the apartheid period, the freedom struggle and political imprisonment in South Africa. These collections have contributed to many notable publications, television documentaries, school textbooks and academic works. They not only hold value as research tools, teaching aids and as crucial evidence for the intellectual development of theories and models but they also contain collective social memory. Consequently, Historical Papers is an accessible hub for human rights research serving civil society as well as scholars.

The collections in Historical Papers are accessible through a variety of finding aids available in the Department and online.

Indexing of Photographs

Historical Papers has a large collection of photographs, which needs to be indexed. The index includes the names of people, places and organisations, as well as events, activities (e.g. Mining) and some concepts (e.g. Influx control).

The index is tailored to this unique collection, and has also been shaped by the enquiries received. For example, a dozen photographs of musicians of different eras and styles were initially indexed (along with a few singers and dancers) under 'Entertainment'. The photographs archivist realised that there was considerable interest in early jazz musicians,

and so added a specific entry for these, and a general entry 'Musicians', which now refers the user to several other entries.

An attempt has been made to see the index through the eyes of a researcher – someone who does not necessarily know the scope of the collection. Thus cross-references to similar events or concepts, et cetera, were added. For example, the entry 'Housing' with its subheadings of 'Informal housing', 'Traditional African housing', and so on, includes cross-references to 'Hostels', 'Mine compounds', 'Squatters', and 'Rural' and 'Urban environments'.

It is sometimes necessary to have several index entries for one photograph. Thus a significant photograph showing the leaders of the Women's March of 1956 at the Union Buildings is indexed under the headings 'Women involved in resistance' and 'Resistance: Passes', and there are also entries for each of the four leaders of the march.

In compiling the index, it was necessary to use the correct terms to describe concepts, according to the thesaurus of accepted terms. For example, the preferred term for forced removals is 'Relocations'; for unrest, the preferred term is 'Violence'. The preferred terms have been used, but have also included entries for the commonly known terms (e.g. Unrest: *see*: Violence), because of the belief that researchers might look for these.

The index is forever growing – it is an ongoing challenge to make it as accessible and 'user-friendly' as possible.

Some Problems with Indexing in the Archives

When a member of the public entrusts to the archives a diary written by a forebear who was present at the Siege of Mafeking, one can see the thought bubble 'This won't disappear for ever into a dusty archive and never see the light of day again … will it?' The donor is reassured that the item will be listed in the *Guide to the Archives* and selected names and subjects will be indexed. But it is in the selection of these names and subjects that the greatest challenge arises – for who is the ultimate judge of what entries should be included in the index? Are the archivists the main determinants of what constitutes history? What is selected for archiving and indexing is what future researchers will use when recording the events of the past – a heavy responsibility.

It is possible to over-index. In the past busy researchers have been seen clicking their tongues with impatience when, having waited for the tenth box to come up from the basement in response to a request for 'any items written by Chief Luthuli', they find that the indexed reference is an acceptance to an invitation for tea with the Bishop. Researchers will not usually appreciate having to spend time on trivial index entries.

But these days there is a greater danger of under-indexing, for two reasons. Firstly, the volume of material that comes into the archives makes it impossible to read and index everything. When minute books from Diocesan Offices come in, one knows they are packed with historical information about churches that someone is going to ask for sooner or later when writing a parish history, but indexing every reference to every church is often too time-consuming a task.

The second reason is that it is difficult for indexers to know what names and subjects are going to be of historical value, either because research requirements change over the years

or because indexers may not have enough in the way of general knowledge to identify the names or subjects that are worthy of inclusion in indexes. Hundreds of people had occasion to write letters to Archbishop Tutu or to Helen Suzman (to name two people whose papers are kept in the archives). Many of them are just ordinary people with no particular claim to historical importance. But some may be people whose views carry weight, or whose letters may be of future significance. Mary Renault? Chester Crocker? Robin Renwick? Lionel Abrahams? If these names do not ring a bell with the indexer their letters are likely to end up in a file marked something like 'Correspondence Jan-Apr 1976'. Then the donors of papers may feel that the archives have let them down.

Fashions change in research. One obvious change is the growing interest in black history: whereas in the past the archivist may have indexed a file under subject matter 'Abortion sub-committee – minutes of meetings', thinking that the contents had been covered, these days people may be asking how many black delegates were present at a meeting and how many times they spoke. (Can this be indexed?) In the old church records, the missionaries would write enthusiastically about the virtues and activities of their black catechists and deacons, sometimes giving a first name only. (Perhaps they were writing reports for the missionary body overseas and thought surnames of catechists irrelevant.) Research would be necessary to ascertain how to index these groups of people, who are currently a popular focus for study.

In order to provide a really detailed account of the contents of a set of correspondence one would need to do a chronological 'calendar' of papers – as was done for A. B. Xuma – with the date of the letter, name of the correspondent and summary of the contents – and then index the calendar by names of correspondents and subject matter. This is quite labour-intensive and can only be done with a few important collections. Nowadays archivists tend to put correspondence into files according to subject matter or date, and index names and subjects that seem 'significant'. It is useful to have the same people indexing the records and helping researchers – in this way one gets a feel for required information. It is known that church records are scanned for individual churchmen, histories of parish churches and missions, church policy and statements on political and social issues. Archivists may be taken by surprise when users request material on some bizarre subject – they may have a vague recollection of having seen something relevant but it may never have been indexed because it did not seem likely as a subject of interest for research.

Conclusion

The qualities needed in an inspired archival indexer will include (as well as careful attention to detail, and a wide general knowledge) the kind of imagination that will predict future research requirements and supply those useful keywords that will turn a folder of dull-looking documents into an exciting treasure trove for researchers.

20 EDUCATION

Sophia le Roux

Abstract

This chapter argues that in order to index educational material effectively, the indexer needs to be thoroughly acquainted with the workings and trends of the education system and, in the case of curriculum material, be familiar with the terminology of the national curriculum and the particular subject field. It sketches the predominant changes affecting the education system in South Africa during the past decade characterised by a radical transformation process. It also examines the intellectual skills of prior knowledge, judgment, reading and analysis when indexing educational material. It discusses how the changing landscape of education impacts on the indexing of educational material, concluding that keeping abreast of the changes in education and the needs of the end users will contribute to better access to educational material by potential users.

Introduction

During the first decade of democracy, the South African education system underwent drastic changes in transforming itself into a nonracial, nonsexist, democratic and equitable system. The purpose of this chapter is to examine how these changes impact on the indexing of educational material. It is argued that the indexer of educational material needs to be thoroughly acquainted with the education landscape in order to index effectively, so that educational material can be fully accessed by the end users. Knowledge of the subject field and the users of this material and also possession of intellectual skills such as the ability to exercise sound judgment, and to do careful reading and thorough analysis of the text are all indispensable to effective indexing.

The Transformation of Education in South Africa

Following the General Election on 27 April 1994, a new constitutional dispensation came into being in South Africa. In terms of the *Constitution of the Republic of South Africa Act 200 of 1996*, South Africa was divided into nine provinces. The first task of the new national Department of Education (DoE) was to transform the fourteen racially and ethnically based executive education departments into nine integrated executive provincial education departments (South Africa 1997:7). The DoE is responsible for formulating policy, setting

norms and standards and monitoring and evaluating all levels of the education system. This comprises 32 910 public institutions ranging from early childhood development centres to higher education institutions (South Africa 2003:1). To transform the curriculum, in October 1997 the Minister of Education introduced *Curriculum 2005* (*C2005*) as the new outcomes-based National Curriculum Framework for General Education and Training (school education). In 2002, after experiencing many problems in implementing this curriculum, *The Revised National Curriculum Statement for Grades R-9 (Schools)* was approved by the Cabinet and the Council of Education Ministers for General Education. In 2003 the *National Curriculum Statement Grades 10-12 (Schools)* for Further Education and Training was introduced. The establishment of the National Qualifications Framework (NQF), the promulgation of the *South African Qualifications Authority Act 58 of 1995* and the enactment of the *South African Qualifications Authority* (SAQA) were all geared to opening up learning opportunities and career progression for all South Africans.

The DoE also embarked on the restructuring of higher education and further education and training institutions. This process reduced the former 152 education and training colleges to 50 mega-institutions. Higher education institutions were reduced from 35 to 24 through a process of mergers and incorporations (South Africa 2003:i,10). The financing, resourcing and cost of education in public schools are other issues high on the agenda of the DoE in its attempts to improve accessibility of education to the poor of the country (South Africa 2003:i). Other priority policy issues include teacher education and development and, in particular, the empowerment of the teaching staff at schools to implement the new curriculum effectively in the classroom; the quality of the teaching and learning of mathematics, science and technology in the schools; the improvement of the literacy rate; national examinations and assessment; HIV/Aids; and the performance of schools. Collaboration between the DoE and the Department of Labour is evident in shared programmes such as the Human Resources Development (HRD) Strategy for South Africa and skills development which includes learnerships.

Categories of Indexing

Indexing is about making information findable (MacDougall 1996:282). It entails the examining of documents, determining their subjects, and selecting appropriate indexing terms reflecting those subjects and their essential meaning. Indexing of educational material falls into three categories: book or 'back of book', and serial and database indexing.

Book indexing is aimed at providing in-depth access to the contents of an individual book and is mainly based on the author's own terminology. Serial indexing entails the ongoing indexing of serial titles. For the sake of consistency over time, the development of a standardised terminology, such as a thesaurus, is preferable. Database indexing is used for online retrieval and is performed by indexing and abstracting services as well as by information professionals of various organisations, such as academic institutions and government departments. It covers a wide variety of materials including journal articles, reports, conference papers and books. With database indexing a controlled vocabulary, in the form of a thesaurus, is preferable to the use of natural language.

Intellectual Skills Needed by the Indexer

For the effective indexing of educational material the indexer should be endowed with certain intellectual skills. Ward (1996:217-223) identifies several of these skills involved in effective indexing which are also applicable to indexing of educational material. *Prior knowledge* of the subject field and of the users of educational material is of crucial importance. To index effectively, the indexer needs a thorough knowledge of the aims and policies of the DoE, the functions and activities of the executive provincial departments of education and the methodology of the national curriculum. Since the new democratic government came into being, a wide range of new education terms has developed. These terms are already embedded in the material acquired by libraries and information centres. The indexer needs to be familiar with the latest education terminology and to keep abreast of any changes and trends in this field. Finally, the indexer has to be familiar with the general and specific needs of the users of educational material.

Judgment while handling educational material is important, as the indexer has to make decisions about the document and its approach, for example decide what to index and the degree of exhaustivity to apply to each document. This could affect the use of a document, as it is sometimes not until a document is being indexed that its full significance and usefulness for the target audience is determined. *Reading* the text, according to Ward (1996:218), means 'mentally incorporating the structure and content of a text not in and for itself, but as it is of value to the user'. The indexer has to read the text as if he or she were the end user of the text or text base. Ward (1996:218) argues that the indexer therefore must have adopted something of the stance, inclinations and presuppositions of the users of the material, and that such empathy can only be acquired through personal knowledge of, and identification with, the end users.

Careful analysis is imperative. This entails three main processes, each involving an act of comparison (Ward 1996:220-221). The first is the comparison of each text with all the other texts that have preceded it in order to identify new features worth indexing. The second comparison is between each text and the indexer's personal experience. Does the text contain new concepts which the indexer has not encountered before? In that case these terms have to be recognised and explained in a scope note in the thesaurus. The third act of comparison is that between the text and the users' known needs and expectations. Which information in the text is relevant to the users and has to be highlighted? Is there crucial information hidden somewhere in a lengthy report?

Indexing Educational Material

With book indexing knowledge about the subject field and its terminology is of the utmost importance while the target audience must be continually taken into account. Is the book a textbook for school learners, for students or for educators? By identifying with the users of the index, the indexer will ensure that the index terms are appropriate to the retrieval of the needed information. When indexing learner support material indexers should familiarise themselves with the official terminology of the particular learning area. This can be obtained from the curriculum and assessment glossary at the back of each of the eight learning area

statements which form part of *The Revised National Curriculum Statement for Grades R-9 (Schools)* and the *National Curriculum Statement Grades 10-12 (Schools*) issued by the DoE. This will facilitate the making of cross-references in the book index, directing users from recognised terms not chosen for index headings to headings used and thus ensuring standardisation.

With serial indexing as well as database indexing, a thorough knowledge of the trends in education is required. A good starting point to acquainting oneself with these trends and the official terminology is to study the legislative framework within which the education system is built. The relevant legislation, such as the *National Education Policy Act, 1996* and *The South African Schools Act, 1996 (SASA),* all give a clear indication of trends in education and also provide glossaries which contain the official terminology. Policy documents and departmental reports, such as *The Review of the Financing, Resourcing and Costs of Education in Public Schools* (South Africa 2003:9) and annual reports, are primary sources to assimilate the recognised terminology used by the education authorities. These can be used by the indexer to develop an in-house thesaurus and may also provide definitions of the terms which can be utilised by the indexer for the scope note of the relevant term in the thesaurus. Most of these documents are available on the Internet and can be directly downloaded from the websites of the DoE and provincial education departments.

One problem often confronting the indexer is the proliferation in the use of acronyms for both names of organisations in the education field and education concepts. When developing a thesaurus it is useful to include a component (field) called the *identifier.* Identifiers are usually the proper names of governmental and nongovernmental bodies, names of Acts and Bills and names of concepts, such as LSEN for 'Learners with Special Education Needs'. The full name of organisations and concepts can be recorded in the identifier field with the acronym in brackets after the name. Identifiers are also used in the *Thesaurus of ERIC descriptors*, developed under the auspices of the Educational Research Information Center (ERIC) of the Office of Educational Research and Improvement of the United States Department of Education. This thesaurus is also an excellent guide for indexers of educational material in South Africa.

In the case of educational material dealing with curriculum matters, it is important for the indexer to keep abreast of the changes in terminology. From the implementation of *C2005* to the introduction of *The Revised National Curriculum Statement for Grades R-9 (Schools)*, many curriculum terms have changed, such as the term 'critical cross-field outcomes' which has become 'critical outcomes'. The indexer must be aware of these changes and make cross-references between the old and the new terms. New terms have also been adopted, such as 'developmental outcomes', of which the indexer must be aware when indexing curriculum material.

Conclusion

In this chapter it has been argued that effective indexing of educational material can only take place if indexers of such material keep abreast of the latest developments in the education field and know the particular needs of the end users of this material. Only then can the aims of indexing be achieved – of bringing the contents of documents to the attention of

potential users. This requires commitment by the indexer as well as the intellectual skills of prior knowledge, judgment, proper reading and analysis of the text.

BIBLIOGRAPHY

MacDougall, S. 1996. Rethinking indexing: impact of the Internet. *The Australian Library Journal*, 45(4):281-285.

South Africa. Department of Education. 1997. *Annual report 1996-1997.* Pretoria: Department of Education.

South Africa. Department of Education. 2003. *Annual report 2002-2003.* Pretoria: Department of Education. Also available at: http://education.pwv.gov.za/index.asp?src=docu&xsrc=repo (accessed 1 May 2004).

Thesaurus of ERIC descriptors. 1995. Edited by JE Houston. 13th ed. Phoenix, Ariz: Oryx.

Ward, ML. 1996. The future of the human indexer. *Journal of Librarianship and Information Science*, 28(4):217-225.

21 ENVIRONMENTAL ISSUES

Martha Pretorius

Abstract

Indexing of environmental material requires special skills and knowledge. This chapter provides guidelines for the indexing of publications on the wider environment, focusing on chemical, geographical, physical and social features.

Definition of the environment: 'the sum total of all biological, chemical and physical factors to which organisms are exposed' (Deason 1986).

Introduction

People have been inevitably linked to the natural environment from the beginning of human time. The first farmers learnt when to plant and when to reap their crops, ancient settlements were built close to safe and reliable water sources, while women had to learn which plants were suitable for cooking and which were better left alone. It was, however, only during the latter half of the twentieth century that the role played by humans in environmental damage and destruction was fully appreciated. Today, most of us are familiar with environmental terms such as 'global warming', 'environmental impact', 'the greenhouse effect', 'biodiversity' and 'sustainable development'. As more and more environmentally related scientific publications see the light, it has become essential to index environmental literature to ensure easy and effective retrieval of information as required to best manage and control the environment. This chapter provides guidelines on the indexing of environmental publications for the building and development of bibliographic databases.

Historical Background

In South Africa there have been many attempts to order and classify material on a wide range of environmental topics. A report on South African databases published in 1991 (McGillvray, Op't Hoff & Schnugh 1991) revealed that of the 357 databases listed, no less than 110 (or 30,8%) covered numerical or textual material on environmental subjects. Many of these databases no longer exist. Of those still in existence the most prominent is WATERLIT, one of the oldest and certainly most comprehensive locally developed databases on an environmental topic.

The South African Water Information Centre (SAWIC) began the production and development of the WATERLIT bibliographic database in 1975 with funding provided by the Water Research Commission (WRC) and technical support from the Council for Scientific

and Industrial Research (CSIR). SAWIC was disbanded at the end of 1996 although the CSIR continued to develop and maintain the WATERLIT database. In 1998, a decision was made for WATERLIT to be managed exclusively by the Water Research Commission. Agreements were negotiated with various institutions countrywide that would make their unique journal and publication collections available for indexing and to be included in the WATERLIT database. As from 1 April 2003, the National Inquiry Services Centre (NISC) in Grahamstown on behalf of the Water Research Commission now handles production, development and maintenance of the WATERLIT database. The guidelines for the indexing of material on environmental issues discussed in this chapter are based on the guidelines applied in the construction and development of the WATERLIT database (Pretorius 1983; Shelwell 1999).

Basic Indexing Principles

General indexing principles are applied in the indexing of material on environmental issues. Keywords, consisting of single words or phrases, are selected (preferably from a controlled vocabulary list) to represent the contents of the publication. Relevant keywords are selected according to the following general criteria:

- Specific meaning. All the keywords that are essential to convey the meaning of the document have to be listed in the index, whether they appear in the text or not. Many of these terms, or synonyms for them, are usually found in the title and abstract of the document.
- Ease of retrieval. The keywords used to describe the contents of the publication should be selected to ensure that the end user will be able to retrieve that particular publication in the easiest and most direct way. Indexers should ask themselves if anyone would be specifically interested in the information covered by this publication and which terms will be most likely used as search terms to enable retrieval of this information. Therefore keywords should always include at least one main broad term covering the main topic with which the publication is concerned. For example:

> When selecting keywords for a publication on the **design of filters**, the indexer has to indicate where such filters will be used, that is in the purification of industrial effluents, the removal of dust particles from air, treatment of water for domestic use, et cetera. Keywords like 'industrial wastewater treatment', 'air pollution' or 'water treatment' should be selected to indicate the main focus of the publication.

Basic Concepts

Four basic concepts can be identified in any document or publication:

- Why: aim or purpose of the activity, operation or process
- What: item or object in relation to which the activity, operation or process occurs
- How: means of achievement (e.g. techniques, methods, apparatus)
- Where: environment (e.g. geographical location, specific industry, solvent)

Keywords

The number of keywords selected for the indexing of a single document will depend on the database policy. Superficial indexing may require a maximum of 5 keywords, while more in-depth indexing may require up to 25 keywords. Be careful not to over-index any item – keep in mind that a *set* of keywords will be used to retrieve information from the database.

Indexing Requirements

Because of the scientific nature of most publications on environmental topics, it is recommended that indexers used for the processing of environmental material should have at least a first degree or higher diploma (awarded by a technikon or technical university) in the natural sciences. The ideal would be to use an indexer with a qualification in botany for publications on vegetation, a chemist for the indexing of publications on environmental chemistry, and so on.

Vocabulary Control

Keywords selected to describe the contents of a publication fall into two categories:

- Descriptors – words or phrases from a controlled list of terms.
- Identifiers – free-language terms. Identifiers may include geographical names, trade names, names of procedures, processes or techniques, names of chemical compounds, species names of organisms, et cetera.

For more detailed information on vocabulary control, consult chapter 14 on thesaurus construction.

Spelling

Avoid the use of American spelling in the allocation of keywords, especially when selecting identifiers. American spellings should be changed to their standard South African variants. For example:

Aluminum	use	Alumin**ium**
Cesium	use	C**ae**sium
Color	use	Col**our**
Odor	use	Od**our**
Sulfur	use	Sul**ph**ur (also sul**ph**uric acid, sul**ph**ates, etc.)

However, species names must not be altered:
Desulfovibrio desulfuricans, Desulfobacter, et cetera

Indexing of Environment-specific Information

Analytical Chemistry

Indexing publications on analytical chemistry calls for great care in the selection of keywords. Applying the four basic concepts listed above (why, what, how, where), indexers

may have problems identifying the key issues in reports on closely related, but totally different, chemical issues. The following examples explain the difference:

Example 1:
Title: Determination of polycylic aromatic compounds in oysters using high performance liquid chromatography.
This article is mainly concerned with the **analytical technique** applied in the determination of the polycyclic compounds.
Keywords: **analytical techniques**; animal tissue analysis; oysters; high performance liquid chromatography; polycyclic compounds; ....

Example 2:
Title: Identification of polyhalogenated anisoles and phenols in oysters collected from Table Bay.
This article reports on the **results** of a survey to determine the effects of pollution on oysters collected in Table Bay.
Keywords: **pollutants identification**; oysters; animal tissue analysis; phenols; anisoles; gas chromatography; Table Bay; ....

In Example 1, full particulars are given of the sample preparation, the equipment used, any enrichment techniques applied, interfering substances, the concentration range for which the method is applicable, and so on. As a final sentence in the publication, reference might be made to the fact that this particular technique has been used in the determination of those compounds in oysters collected from Table Bay.

In Example 2, only limited attention is paid to the analytical technique used, but the *results* of the analysis will be reported in detail.

Chemical Names

The full names of key chemical compounds mentioned in the text need to be included as identifiers (unlisted keywords), especially when the title only provides the chemical formula of such compounds:

Title: The Occurrence of As, Cd and Co in Soil Samples from the Limpopo Province in South Africa.
Keywords: soil contamination; Arsenic; Cadmium; Cobalt; ....

Heavy Metals

As mentioned before, it is essential to select a broad term as keyword to identify the main application field of the research reported in the publication, for example type of organism (aquatic insects, soil bacteria, marine plants), environmental conditions (cold regions, arid climates) and problem areas (waste disposal sites, informal settlements). As a group of chemical compounds, heavy metals and the detrimental effect that they may have on the whole environment are of great concern to environmental chemists. It is therefore useful to use 'heavy metals' as a broad term whenever any one of the greater group of heavy metals (metals whose specific gravity is approximately 5 or higher) is discussed.

Greek Letters

Any Greek letters appearing in the title need to be written out in full when used as identifiers:

> Title: Making waste sterile - with γ-**rays**
> Keywords: radioactive waste disposal; radioactivity; **gamma rays**; ....

Indexing of Publications in the Biological Sciences

Species and Family Names

Publications on the biological sciences will most probably include the name(s) of a specific species. The species name, as well as the broader genus, family or order names relating to that specific animal, plant, organism, et cetera, should be added as an identifier:

> Title: The effect of some ecological factors on longitudinal patterns of black fly community structure in a foothill stream.
> Keywords: diptera; simuliidae; blackflies; aquatic insects; .....

Any species name appearing in abbreviated form in the title of a publication needs to be included in full as identifier:

> Title: High levels of **E. coli** detected in food samples following a sewer burst in an urban township.
> Keywords: food contamination; **Escherichia coli**; ....

Scientific and Popular Names

Scientific names as well as popular names of fauna and flora should be used as identifiers:

> Title: Lower numbers of Lesser and Greater Flamingos migrating to drought-stricken Botswana.
> Keywords: Lesser Flamingo; **Phoenicopterus minor**; Greater Flamingo; **Phoenicopterus ruber**; Botswana; droughts; ....

Abbreviations and Acronyms

Any abbreviations in the title referring to processes, constants or organisations need to be included in full as keywords:

Example 1:
Title: Determination of pollutants in water by **GC/MS**
Keywords: pollutants determination; water analysis; **gas chromatography-mass spectrometry**; ....

Example 2:
Title: Estimating the **LC 50** values in experiments determining the toxicity of lead to Rainbow Salmon.
Keywords: toxicity tests; rainbow salmon; salmo gairdneri; lead; **median lethal concentration**; ....

Example 3:
Title: Report on the role of the **WHO** in the control of human diseases in Africa.
Keywords: human diseases; disease prevention; Africa; **World Health Organization**; ....

Depending on database development policy, standardised acronyms for diseases, company names, pesticides, et cetera, may be used as legitimate keywords:

- HIV/Aids (disease)
- 2,4 D, TDE, DDT, EPTC, MSMA, MCPA (pesticides)
- SASOL, ESKOM (company names)
- pH (a measure of the acidity or alkalinity of a solution)

Geographical Names

In many publications on the environment, details about the geographical location of the reported event are of the utmost importance. The effect of atmospheric pollutants released by an industrial plant located in a densely populated area will obviously be more severe, compared to the effect the same event may have where the same industrial plant is located in a remote area. It is therefore important to include geographical references as identifiers, for example country or regional names, names of towns or cities, rivers, dams, lakes, seas and oceans, mountain ranges and conservation areas. It is recommended that standardised lists of important geographical names should be developed for this purpose, either by the database owner, or at least by the indexer. When geographical names are used in their non-English form, the standard English equivalent should be used as keyword. For example:

Wien	use	Vienna
Bodensee	use	Lake Constance
Firenze	use	Florence

South African geographical names are obviously of great importance to local users. Indexers should carefully select geographical identifiers, paying special attention to spelling and recent name changes (provinces, towns, dams, rivers, etc.).

Social Issues

During the past few years, the impact of environmental conditions on the social characteristics of communities has received great attention from researchers, governments and international institutions. Human poverty and other socioeconomic problems are often the result of existing environmental conditions such as droughts, over-exploitation of natural resources and natural disasters. Indexing of material on these issues needs to include terms on the environment as well as on the social or socioeconomic content of the publication.

Conclusion

The first World Summit to discuss global environmental issues was held in 1992 in Rio de Janeiro, Brazil, followed by the Earth Summit which took place in Johannesburg, South Africa, in 2002. The outcomes of these two events, and the targets set for future and immediate actions aimed at improving environmental conditions for all the earth's inhabitants, clearly indicate that environmental issues will remain a global priority well into the twenty-first century. It is therefore of the utmost importance, also for all interested parties in South Africa, to keep abreast of research and new developments in the environmental area. Proper indexing of environmental material will help achieve this goal.

BIBLIOGRAPHY

Deason, WO. 1986. *Environmental glossary*. Denver, Colo: United States, Department of the Interior, Bureau of Reclamation.

McGillivray, R, Op't Hoff, M & Schnugh, L. 1991. *Directory of South African databases*. Pretoria: Division of Information Sciences, CSIR.

Pretorius, M. 1983. *Indexing for the Waterlit Data Base – a practical manual*. Pretoria: SA Water Information Centre, CSIR.

Shelwell, J. 1999. *Policy guidelines manual for the Waterlit Database, Volume II*. Pretoria: Water Research Commission.

United States. Department of the Interior. 1986. *Environmental glossary*.

22 HEALTH SCIENCES

Elizabeth Robertson

Abstract

A brief overview of information resources indexed for retrieval in the health sciences looks at needs and opportunities for indexing in the field. Suggestions are made about indexing aids and problems, and the help obtainable from large libraries and online indexes and catalogues.

Introduction

Health sciences indexing is essentially no different from any other subject indexing but is made more complex by the very structured nature of the health sciences and the perceived complexity of medical terminology. Some of the major indexing institutions and their products are considered below.

The National Library of Medicine, Washington

The National Library of Medicine was set up in the 1960s. Besides collecting and preserving the literature of medicine, its mission was to index and make available the content of medical literature. This was done by the publication of *Index medicus,* monthly and with annual cumulations, for up-to-date access to medical literature in serial form worldwide. *The National Library of Medicine catalog* lists monographs and non-serial medical literature. Both appeared in print form until the end of the 1990s, and are now available online.

Two important tools accompany *Index medicus.* They are the *List of journals indexed* (updated annually and now numbering 4 600 titles) and their standard abbreviations. The second is *MEdical Subject Headings (MeSH),* a list of 22 000 descriptors and 24 000 *see* references which will be further discussed below.

Medline and PubMed

Medline is available online as *PubMed* on the Entrez link at the National Library of Medicine website http://www.ncbi.nlm.nih.gov/ – PubMed contains the whole *Index Medicus* database and subsidiary indexes, and *MeSH* and the *List of journals indexed.* It is available free of charge, and *MeSH* and the journals list can be downloaded. Since they are extensive, downloading can be slow.

Google

Much further information can be found, more informally, on the Internet. Google is the serious researcher's search engine of choice.

ISAP

The *Index to South African periodicals (ISAP),* accessed online through SABINET or online or on compact disc from *NISC* databases *African studies* and *South African studies,* provides access to the contents of South African journals of medicine and the allied health sciences (as well as to most other important South African journals). Only seven South African journals are indexed in *Medline.* This includes two veterinary journals, but no nursing journals. Indexing of the *South African Medical Journal* in *Medline* is incomplete, since many valuable sections are issued separately and do not appear in *Medline.* It is important to know therefore that *ISAP* is more comprehensive in the indexing of this key journal and of South African journals in general.

Book Indexing, Personal Files and Small Collections

The above indexes seem so comprehensive that one might think no further indexing in the health sciences would be required in South Africa. However, published books require indexing and book indexers need to follow standard indexing conventions. NGOs still need to index their in-house resources, some of which include the health sciences. And academic researchers collect personal information in all formats, which need organising. All of these benefit from the use of appropriate software packages and knowledge of the aids to health sciences indexing.

Aids to Health Sciences Indexing

A comprehensive explanatory medical dictionary is essential for indexing. Many are available. *Mosby's medical, nursing and allied health dictionary* (1988) is designed, as its title indicates, for use by practitioners in allied branches of health sciences and therefore is likely to be helpful for the lay indexer. It is worth noting that most English language medical dictionaries are of American origin and will spell 'paediatric', 'haemolytic' or 'foetus' as 'pediatric', 'hemolytic' and 'fetus'. It is therefore useful to have a small standard medical dictionary of British origin and a good general English dictionary to check standard English spelling.

It is also interesting to note that medicine, in the literature at least, and in health science faculties, is becoming more democratic. The comprehensive term 'health sciences' is used, and the contributions of nursing, physiotherapy, occupational therapy, pharmacy and other allied professions are acknowledged and valued.

The Major Indexes

The major indexes of the health sciences are *Index medicus*, *Excerpta medica* (known online as *Embase*), and the *Comprehensive index to nursing and allied health literature (CINAHL).* Other science databases exist which may have health sciences coverage. Their names, scope and accessibility change, as do subscriptions to these expensive tools, and one should seek information about them at large libraries with internet facilities and online subscriptions.

Medline is of great interest because of its long history and its structure. *Index medicus* published its final printed issue of *Cumulated index medicus* in 2000 after 34 years of punctual updating. From 1970 it has also been available online, as *Medline*, in various developing formats, gradually becoming easier to search. It still retains, online, the *List of journals indexed* and *MeSH* which has made it so valuable bibliographically over the years.

List of Journals Indexed

The list of journals indexed, with standard journal abbreviations for citation purposes, is available in *Medline.* Most medical journals require that these standard abbreviations be used, which sometimes presents an author with the problem of how to abbreviate non-medical journals which may also be cited in the same work. The 'Vancouver style' was adopted in 1982 by a conference of medical editors, a large group of whom agreed to use a standard form in the publication requirements for authors in their journals. This included the standard journal abbreviations mentioned above and a citation style based on that evolved by *Index medicus.* It is the least fussy and the easiest to use and edit of commonly used editorial styles. The style is detailed in the journals which support it: it appears, for instance, at the beginning of each issue of the *British Medical Journal* and in the *South African Medical Journal.*

MeSH

The major tool of *Medline* is *Medical Subject Headings*, or *MeSH.* There are 22 000 descriptors. Health sciences, from cytology to psychology through a large number of specialisms, is a very broad field. *MeSH* consists of the terms selected in alphabetical order, with *see* references, scope notes and numbers. For example:

ALCOHOLIC INTOXICATION
C21.613.53+ F3.900.100.300+
Human and animal; drunkenness goes here; acute alcohol intoxication = ETHANOL/ poisoning.

These scope notes are instructions to indexers and guidelines for searchers. The existence of such notes gives immediate reassurance. An indexer who wishes to index or find material on the kidney and its diseases, for instance, may wonder whether to look for headings starting with Kidney, Renal, or Nephritic, and *MeSH* provides the necessary guidance.

MeSH is complemented by a further list, a framework known as the Tree Structures, of which the major divisions or trees number about 15, starting with A. Anatomy, B. Organisms, C. Diseases, D. Drugs. These structures form a classification system which contains all the medical subject headings used, numbered and divided by subject down to seven levels of specificity in some cases. It differs from a book classification scheme like the *Dewey decimal classification,* for instance, in that it does not attempt to classify every possible subject. It has literary warrant – that is, it is built to classify the subject matter it indexes and it acts as a framework for parts or divisions of health sciences literature.

MeSH is not a dictionary, but one can learn something about a subject by referring to the tree numbers assigned. The examples below illustrate this:

- The term **Polymyalgia Rheumatica** for instance, shows no comment at all, but a number: C5.651.742. This number, in the trees, shows that the broader term is Neuromuscular Disease and, at the same level of indentation it can be seen that it is related to Arteritis [sic]. Now, without previous knowledge, the medical context of the disease is clear. A medical dictionary will give more detailed information but possibly less context.
- Similarly, under **Cancer** there are only two headings. The see reference, however, indicates that the preferred term is **Neoplasms**, and in the trees under C4 Neoplasms there are 20 pages of classification of every form of cancer.
- There are difficulties about indexing by assigned terms. There may be every reason *not* to use these in a book index, but they do clarify the context in an area where the endocrinologist and the psychiatrist may talk virtually different languages. A researcher may not habitually use a particular term in a subject area but would hope to find see references covering at least some of the terminology of the literature. However, for a long while it was difficult to find articles about stroke in Index medicus, one had to be sufficiently knowledgeable to search under **Cerebral Accident**. (Latterly, a see reference has been added.)
- **Autoregulation** has a *see* reference to **Homeostasis G7.621+**. The + sign indicates that there are further indented (narrower) terms in the tree structures at this number.
- Terminology changes as new subject areas open up. Old headings become inadequate, and appear so. There is a new heading **Nevirapine**, described as follows:

> A potent, non-nucleoside reverse transcriptase inhibitor used in combination with nucleoside analogues for treatment of HIV infection and AIDS. Previous indexing: Azepines (1991-92)
>
> Pyridines (1993-97).

When searching online most possible terms can be retrieved directly, as old headings and new are searched, as well as terms referred from, and all text words used in titles of articles. The fact that *Medline* is indexed by *MeSH* ensures a more powerful and complete retrieval of subject matter than text words alone allow, because of the framework of the trees. Subheadings are available for refining the search, as are dates, indications of format (e.g. a review article) and whether the articles are available online in full text without charge. The accessibility of information has increased exponentially.

Recent Developments in Medical Indexing

Using Medline to search the literature under Abstracts and Indexing, it appears that little has been published on the subject in recent journal literature anywhere. The articles retrieved are about experiments on the efficacy of retrieval using various methods, or the availability on the Internet of small journals excluded by *Index medicus* policy from *Medline* but linked on the Web instead to major websites. The seminal articles on indexing and retrieval and

the majority of manuals and instruction booklets happened in the early 1970s when *Medline* was new. Now that databases are much more user-friendly, having extensive instructions and illustrations in the Help index, librarians are often assumed to be capable of effective searching, and also of effective teaching of end users. If they understand *MeSH*, their results should be more effective than those of some of their clients. The fact that *PubMed* is available online and free of charge makes it possible for every medical indexer to work to internationally accepted standards.

BIBLIOGRAPHY

Mosby's medical, nursing and allied health dictionary. 1998. 5th ed. Edited by KN Andrews. St Louis, Miss: Mosby.

23 LABOUR

Judith Shier

Abstract

This chapter provides an overview of literature published in the field of labour (many without indexes). Indexing needs and requirements are discussed. It also includes guidelines for indexers. A very useful bibliography contains references to many sources that may be consulted for information in the field of labour.

Historical Overview

South Africa has a fairly long, complex and rich labour history that is well documented. The first forms of industrial organisation were craft unions created by skilled immigrants (before Union). Following a period of intense confrontation between organised skilled white workers, the government and employers in the 1920s, an industrial conciliation framework was created. This incorporated white workers and excluded Africans. This framework was challenged by an emerging African working class in the 1940s and 1950s. Severe political repression in the 1950s and 1960s ensured that African workers remained disorganised and outside of the state's industrial conciliation mechanisms. It was only in the 1970s that this framework was challenged and rendered unworkable.

This section reflects writing about the labour movement and industrial relations since the re-emergence of the independent trade union movement in the early 1970s. A relatively small, reasonably well-defined field in South Africa has been generated since then, covering labour legislation, the economy – national and international – workplace issues and politics, as well as aspects of the labour movement itself such as strikes and collective bargaining issues. Approaches range from the highly theoretical or polemical to the more practical and popular. The types of material include 'grey literature', books, journals or magazines, photographs or posters, papers and online material. The trade union movement has played a key role in modern South African history and has been a topic for debate among academics and entrepreneurs, political activists, officials and members of the individual unions and federations. In the struggle for freedom in the workplace and in the economy and society the union movement itself developed a tradition of writing and publishing about political and economic and workplace issues. Much of the work produced from within the trade unions had an educational or campaigning purpose, and had the appearance of 'struggle literature' (especially in the 1980s). Academic institutions also produced scholarly works about the labour movement and labour issues.

Many of the publications in the labour field have been published without indexes. This includes internal trade union material such as 'NUMSA Women Organise' or the Chemical Workers Industrial Union's 'Defend, empower, advance: responses to plant level bargaining'. Papers of the annual Labour Law Conference held in Durban were published from the late 1980s, but not indexed. A long-standing journal on labour issues, *South African Labour Bulletin*, started in 1974, does not have a regular published index, and as a result the full potential of this journal as a research resource is lost. In the 1990s even reputable publishers were publishing books without an index; see for example Standing (1999). A reviewer in the labour field bemoans the fact that the book he is reviewing otherwise quite favourably does not have an index. Of *Twenty years in the labour movement: the urban training project*, published in 1999, he says: 'It's a great pity, though, that a book of about 343 pages does not have an index!' (Buhlungu 2000:71). Even into the 1990s the value of an index was still not fully appreciated.

More recently the Congress of South African Trade Unions (COSATU) website (www.cosatu.org.za) has become a resource for online users, using the Google search engine on its site. When launched this site was accorded the 'Labour Site of the Week' award by an international labour website evaluator. It is a valuable source of documents for researchers and labour activists, and includes the COSATU journal, *The Shopsteward,* which is also searchable via Google. Another recent development is the Digital Imaging South Africa (DISA) project 'Southern African Freedom Struggles 1950-1994'. DISA staff have been responsible for indexing (among many other journals) six labour newspapers such as *Fosatu Workers News*, *NUM News* and *Cosatu News* online. Up to now these have been difficult to locate but they are now available in full text.

Indexing Difficulties, Unique and not so Unique

Audience

Identifying one's audience is an essential prerequisite for an indexer. These are likely to be students and academics, union officials and members, activists, NGOs and employers. Certain material is produced for a union audience alone ('How to cost a wage demand'), while others are designed to be read by anyone involved or interested in industrial relations, for example the *South African Labour Bulletin*, or the annual proceedings of the Labour Law Conference held in Durban since the late 1980s. Some material is scholarly, some political, mainly written in English, but increasingly from the 1970s and 1980s in indigenous languages to overcome workers' lack of proficiency in English. Assessing the potential audience accurately is an essential key to a good index.

Indexing for the Internet particularly raises questions of this nature: Who is the audience – local or international, or both? How does one achieve the consistency or control of keywords evident in a book index? In addition, which education level is one aiming for? Another problem area in a multiple indexing project is skills training and quality assurance. These are necessary elements since not every indexer approaches the area in the same way; consistency across the project needs to be ensured.

Choice of Terms

This is governed by knowledge of the field and its nuances, the text to be analysed for indexing, and its purpose, as well as an understanding of the audience. Questions to ask when choosing terms: How will the material be used? For what purpose and by whom? There are broadly two types of material: on or about labour, as distinct from material written for the labour movement. Much material analysing the economy of South Africa and industrial restructuring was produced at a time (the 1990s) when COSATU was influential in the creation of economic strategy and the Reconstruction and Development Programme (RDP). This type of index requires a combination of conventional economic terms (e.g. investment, inflation, unemployment) and terms unique to the labour movement or broader social movements. (For example: 'Neoliberalism; September Commission of 1997 – Commission headed by Connie September, that is, not the month of the year.' At the time the significance of the 'September' in the 'September Commission' was well known within the labour movement, but may not have been familiar outside of it.) A possible conflict exists between common terms that have an activist flavour, and may also be specific to that time period and terms which will have a more neutral meaning and will last longer and have wider applicability. Labour discourse, because of its political nature, has tended to reflect rapid shifts and changes in meaning. Terminology and rhetoric can change from year to year. A familiarity with current expression and terminology, and a sense of its purpose, is required. Good indexing presupposes more than pure technical skill – knowledge of the field is essential. Jargon and abbreviations which are unfamiliar to a wider audience should be avoided, for example EPZs (export-processing zones) or ESOPS (employee share ownership programmes). Even the familiar acronyms NEPAD and GEAR will lose currency and meaning over time.

Partly due to the politicised nature of the material the choice of terms also depends upon one's orientation. For instance, the term 'capital' may not have the same meaning for union activists and their counterparts in business. The indexer should be aware of bias.

Acronyms, Inconsistencies and the Alternative Press

Before professionalisation of the independent trade unions occurred many of the publications for labour, especially emanating from the unions, were published by the alternative press. This often meant there was no index, and/or inconsistencies with regard to page numbering. In addition, while the material produced by the union movement or its scholars comprises a sizeable chunk of the South African alternative press output, its purpose was to campaign or educate. Supporting scholarly work does not exist in the form of reference books. Reference books to which an indexer can refer are few and far between. See the bibliography for a selection of those which do exist.

There are many different organisations in this field, and acronyms can litter a text. (A current example is a recently published book on South Africa, *State of the nation*, which has four pages of acronyms listed in the preface.) For indexing purposes, reference lists of names of trade unions and other organisations may have to be created independently. This is because material was often directed at the 'converted' reader and so acronyms of organisations or unions were not listed in full in the text as it was assumed they were

known to the reader. On top of this, organisational mergers and changes of name occur with frequency, so tracking the earlier names can be difficult.

In the case of journals, the numbering of issues may be unreliable, and page numbers awkward. Articles may begin on one page and continue several pages later on in the publication. This could create difficulty for the indexer of articles.

Promotion of Indexing in the Mainstream Press or Otherwise

It is surprising, given the awareness of the educational purpose of publications in this field, that indexing has been somewhat overlooked. We look forward to a time when more publications (especially from NGOs and trade unions) include an index.

Practical Guidelines

- Have a well-developed sense and understanding of the overall project and its purpose. This may only be gleaned from reading the text to be indexed. Occasionally there is an opportunity to consult the author.
- Ascertain the possible audience, and have a sense of how the material will be used. The indexer should be guided by the following questions: What will the reader want to know? What is significant in the text? Who is the author addressing? Who is the audience?
- Familiarity with the subject is an advantage in order to comprehend the different levels of possible indexing, the significance of concepts and the nuances of terminology.
- Make use of reference books or key texts available.

BIBLIOGRAPHY

Adler, G & Webster, E. 2000. *Trade unions and democratization in South Africa, 1985-1997.* Johannesburg: University of Witwatersrand Press.

Annual report on employee benefits and labour relations in South Africa. Randburg: Andrew Levy and Associates. (Annual).

Baskin, J. 1991. *Striking back: a history of COSATU.* Johannesburg: Ravan.

Buhlungu, S. 2000. Review of '20 years in the Labour Movement: the Urban Training Project and change in South Africa, 1971-1991' by D Lowry. *South African Labour Bulletin,* 24(4):69-71.

Buhlungu, S. 2003. The state of trade unionism in post-apartheid South Africa, in *State of the nation South Africa 2003-2004*, edited by J Daniel et al. Johannesburg: Human Sciences Research Council.

Chemical Workers Industrial Union. 1997. *Defend, empower, advance: responses to plant level bargaining.* Johannesburg: CWIU.

COSATU. 1987. *Political economy: South Africa in crisis.* Pietermaritzburg: COSATU.

COSATU. http://www.cosatu.org.za/ (accessed 28th April 2004).

DISA. http://disa.nu.ac.za/disaind.htm (accessed 28th April 2004).

Du Toit, B. 1978. *Ukubamba amadolo: worker struggles in the South African textile industry.* London: Onyx.

Freedom from below: the struggle for trade unions in South Africa. 1989. Johannesburg: Skotaville.

Industrial Health Research Group. 1998. *The Occupational Health and Safety Act: an introduction.* Cape Town: IHRG.

Isaacs, S. 2002. *South Africa in the global economy: understanding the challenges, working towards alternatives.* Rev ed. Durban: Trade Union Research Project.

Jarvis, D. 1997. *Reading the financial pages of a newspaper.* Durban: Trade Union Research Project.

Joffe, G. 1995. *Improving manufacturing performance: report of the South African Industrial Strategy Project.* Cape Town: University of Cape Town Press and ISP.

Lloyd, C. 1994. *Work organisation and world class management: a critical guide.* Johannesburg: Red Earth *and* Umanyano.

Lundall, P, Schroeder, I & Young, G. 1984. *Directory of South African trade unions.* Cape Town: South African Labour and Development Research Unit.

Lundie, R. 1990. *A dictionary of acronyms.* Pietermaritzburg: Centre for Adult Education, University of Natal.

Maree, J. 1987. *The independent trade unions, 1974-1984: ten years of the South African Labour Bulletin.* Johannesburg: Ravan.

McCormick, D & Schmitz, H. [2003]. *Manual for value chain research on homeworkers in the garment industry.* New York: Women in Informal Employment Globalizing and Organizing *and* Brighton: Institute of Development Studies, University of Sussex.

Merrett, C. 1993. *Thesaurus of South African socio-political and economic terms from an anti-apartheid perspective.* 3rd ed. Pietermarizburg: University of Natal Library.

Naidoo, R. 2000. *Unions in transition: COSATU in the new millennium.* Johannesburg: COSATU.

National Union of Metalworkers of South Africa. 2001. *Know your demands: booklet for NUMSA shopstewards, 2001-2002.* Johannesburg: NUMSA.

Ncube, D. 1985. *Black trade unions in South Africa.* Johannesburg: Skotaville.

Official South African trade unions directory and industrial relations handbook. Randburg: South African Trade Unions Library. (Annual).

Pityana, SM & Orkin, M. 1992. *Beyond the factory floor: a survey of COSATU shopstewards.* Johannesburg: Ravan.

Proceedings of the Thirteenth Annual Labour Law Conference held in Durban 19th-21st July 2000. Durban: Butterworth.

South African human rights and labour law yearbook. Cape Town: Oxford University Press *with* Durban: Centre for Socio-Legal Studies.

South African Labour Bulletin. Johannesburg: Umanyano.

Standing, G. 1999. *Global labour flexibility: seeking distributive justice.* Basingstoke: Macmillan.

Standing, G, Sender, J & Weeks, J. 1996. *Restructuring the labour market: the South African challenge.* Geneva: International Labour Organization.

Trade Union Research Project. 1995. *Analysing company information.* Durban: Trade Union Research Project.

Trade Union Research Project & Industrial Strategy Project. 1995. *The motor vehicle assembly and components industry in South Africa: towards an industrial strategy.* Durban: TURP. (This forms part of a series of publications on the Chemical, Paper, Household Electronic Durables and Engineering Industries by the same authors.)

Von Holdt, K. 2003. *Transition from below: forging trade unionism and workplace change in South Africa.* Pietermaritzburg: University of Natal Press.

24 LANGUAGES

Khoe-San: A Botswana Experience

Shelagh Willet

Abstract

The chapter indicates problems which may face indexers unfamiliar with these languages and shows how the problems were dealt with in the context of compiling an annotated bibliography on the Khoe and San at the University of Botswana. Gives a brief history of the Basarwa Bibliography Project (now the Khoe-San Bibliography Project). Provides a style and typing guide for the Khoe, Ju, !Ui and Taa languages. Shows that the use of click signs at the commencement of keywords to be used for indexing creates special problems when typed on a computer, explaining how these were approached in The Khoe and San: an annotated bibliography. Also points out the confusion that may exist when identifying which Khoe-San language is being referred to in a text, as there has been a great lack of consistency in their nomenclature. Provides an overview of the earlier and more recent history of the writing of Khoe-San languages. Emphasises the fact that, currently, several initiatives regarding writing in the Khoe-San languages are being undertaken by the communities themselves. This is by means of oral history projects in the countries of Southern Africa where the languages are still spoken.

Introduction

The suggestions which follow, regarding the indexing of Khoe-San languages, result from the practical experience gained while working for several years with the Basarwa Bibliography Project. The word 'Basarwa' is the term used in the Setswana language to describe the San or Bushman people. The reason for the use of the term 'Khoe-San' will be explained later in the chapter. The Project was established in 1993 under the auspices of the National Institute of Development Research and Documentation (NIR) at the University of Botswana. In 2001 the Project was taken over by the University of Botswana Library. The objectives of the Basarwa Bibliography Project were as follows (Monageng in Lane 2001:74):

- To acquire, as far as possible all contemporary published materials relating to the Basarwa of Southern Africa, that is books, reports, manuscripts, theses, policy documents, journal articles as well as unpublished or 'grey literature'.

- To process this material by the use of a computerised database to facilitate easier retrieval. The processing included making short abstracts of each document.
- To publish the abstracts as a series of bibliographies.

By 1995, when a workshop on Basarwa research was held at the University of Botswana, 300 publications had been entered into the database and a draft volume of the bibliography had appeared. It should be noted that the Bibliography Project was established at the same time as a multidisciplinary programme currently designated the University of Botswana/ University of Tromso Collaborative Programme for San/Basarwa Research and Capacity Building, under the auspices of which a number of research initiatives on Khoe-San have been undertaken, such as journal articles, dictionaries and student exchange arrangements. The Collaborative Programme funded the eventual publication of *The Khoe and San: an annotated bibliography*, the first volume of which was published in 2002 and the second in 2003.

What are Khoe-San Languages and Where are They Spoken?

Indexers working with publications dealing with Khoe-San languages are faced with several problems relating to both the choice of indexing terms and ways of expressing them using computers. Some background regarding the nature of these languages may throw light on the difficulties involved. Firstly, they are totally unrelated to the Bantu family of languages and are characterised by the use of unique sounds known as 'clicks'. Where Bantu languages contain clicks it can usually be proved that these have been adopted as a result of contact with speakers of Khoe-San languages. The complexity of the situation regarding these languages is reflected by the following quotation from a document entitled The Penduka Declaration on the Standardization of Ju and Khoe Languages, which reads as follows:

> In the 1930s European researchers made the mistake of thinking that all click languages came from one source. Previous and subsequent research shows there are at least four entirely separate language families. These are the Ju, Khoe, !Ui and Taa language families. Each language family has different and unrelated grammar, word order and vocabulary. There are various San peoples who speak Khoe languages, and there are non-San people who speak Khoe languages including the Nama and the Damara (Penduka Declaration 2001:1).

It is significant that the delegates at the Penduka Workshop held in Namibia in 2001 were themselves speakers of such languages and were from South Africa, Botswana and Namibia. Between them, they represented eight different language groups. The decision to use the term **Khoe-San** in this chapter is based on a request made by Joram|Useb, the representative of the Working Group of Indigenous Minorities in Southern Africa (WIMSA) at a subsequent conference, in which he suggested that the word previously spelt 'Khoesan' should preferably be hyphenated and 'San' spelt with a capital S (Nigel Crawhall, personal communication). An indexer faced with a publication relating to or in a Khoe-San language may find it difficult to know which language is being referred to, as there has in the past been great confusion in the nomenclature of these languages. In fact, the author of the chapter

entitled 'Khoisan languages' in the book *Languages in Botswana: language ecology in Southern Africa* comments that '[t]he names and spellings of different Khoisan Language names reveal the greatest terminological confusion we have encountered in any scholarly context' (Andersson & Janson 1997:112).

A chapter by Yvonne Treis entitled 'Names of Khoisan languages and their variants', which appears in volume 15 of *Research in Khoisan studies*, deals in great detail with this confusing situation and lists eight reasons why the names linguists, anthropologists, geographers and administrative officers have used to designate a particular group bear no resemblance to what they call themselves (Treis 1998:463). Clearly, if linguists trained in language identification and designation have faced so many difficulties in providing correct and acceptable names for the languages dealt with in this chapter, it is going to be even more problematic for the indexer to do so. To add to the complexities surrounding indexing Khoe-San languages is the fact that they are spoken in several different countries and do not share a common orthography. In some cases the click sounds are represented by the signs employed in the International Phonetic Alphabet while in the case of the Naro language spoken in Botswana, the decision taken has been to make use of certain letters from the Roman alphabet, as is done with the clicks used in Xhosa and Zulu where unused letters, q, c and x represent clicks. Sound reasons have been provided for these decisions but the result is a lack of uniformity and the possibility that related languages may have different orthographies.

Choice of Indexing Terms for Khoe-San Languages and Use of Computers to Represent Click Sounds

In order to provide indexing terms or subject keywords for the documents abstracted in *The Khoe and San: an annotated bibliography* it was found necessary to compile a unique thesaurus of terms not available in any standard list of subject headings. Though based upon the Organization for Economic Co-operation and Development (OECD) *Macrothesaurus for information processing in the field of ecology and social development,* it also contains the terms chosen for the languages referred to in this chapter. Information from one of the appendices from the Penduka Workshop follows, listing the currently accepted names of the various language families and giving an indication of how the signs can be typed using a computer.

Style and Typing Guide for Khoe, Ju, !Ui and Taa Languages: Names of Peoples and Languages

Ju Language Family
!Xun (used to be !Xu): the language and the people
Jul'hoansi: the language and the people
Jul'hoan: use only as an adjective
!Kung: the language and the people

Khoe Language Family
Khoekhoegowab: language of the Hai||om, Nama and Damara peoples
Khwedam: language of the Khwe people
Khwe: the people (used to be Kxoe)
||**Anikhwe**: part of the Khwe peoples, mostly in Botswana
Naro: the language and the people
G|ui or **Cgui**: the language and the people
G||ana or **Xgana**: the language and the people

Taa Language Family
!Xoo or **!Xoon**: the language and the people

!Ui Language Family
N|u: the language of the southern Kalahari
N||n÷e: speakers of N|u
÷**Homa**ni or ÷**Khomani**: people who used to speak N|u and |'Auo
|'**Auo**: extinct language
|'**Auni**: speakers of |'Auo
|**Xam**: extinct language. The San Cultural Centre is called !Khwa ttu. This is |Xam for 'water pan'.
||**Xegwi**: extinct language

How to type special letters: Pull down the Insert menu, choose Symbol, go to Normal, and select the appropriate symbol. | and | | are now on most keyboards, usually in upper case only.

Ordering of Signs and Their Relation to Alphabetisation of Terms

For the compilation of the Subject Index of the bibliography, a specific ordering of clicks was employed as follows:

!
||
|'
|
=

Note the use of | and || in place of / and // which were employed when the thesaurus was devised and which appear in the current Index of the bibliography. A computer does not recognise International Phonetic Alphabet signs and arranges terms containing them as if the first recognisable letter was the beginning of the word. As a result, to avoid this it was decided to list all Khoe-San languages names beginning with click sounds at the beginning of the Subject Index and then to follow the normal alphabetical arrangement of terms. Within each click sequence, the terms were arranged by the first conventional letter following the click. For example:

!Kung,
!Ora,
!Xoo,
!Xun
=Hua,
= K'au//'ei,
=Khomani,
=Kx'au//'ei

In the Author Index the names of authors whose names began with a click sound were placed first and follow the same arrangement. For example:

||Garoeb, I. F. H.
| Useb, Joram=Oma, Kxao Moses.

Within the text of an abstract, the spelling employed by the author was used, even if it differed from the one chosen for the keywords and subsequent use as an indexing term. Thus the author might spell the language '!Xu' and that term would appear in the abstract, but would be shown as **!Xun** in the subject keyword list and Index.

As many of the Khoe-San languages have not yet been written down or provided with an orthography, it is likely that there will be changes in the list of languages in the future because it is the communities themselves who are now involved with capturing the sounds of their own languages and with deciding how they should be written. A few brief details about current oral history projects may help to emphasise the fact that although several San languages are actually in danger of dying out, others are being revitalised through dedicated efforts of educated, literate speakers.

The Preservation of Oral History and Folklore in Written Form

The Early History

The Jagger Library in the University of Cape Town contains a unique archive of written records and translations in the language of the |Xam Bushmen. This priceless heritage was made possible by the dedicated work of the German philologist and librarian Wilhelm H. I. Bleek, his sister-in-law Lucy Lloyd and a band of !Xam speakers who were released from prison to assist them in learning their language. Bleek died at 48 years of age but his younger daughter Dorothea continued the work of her father and aunt and became an authority on the languages of the various San groups as well as editing much of the unpublished folklore that had been collected in the nineteenth century. One of her major achievements was the compilation of a Bushman dictionary which her father began in the 1870s. It was Bleek who first devised a method of writing the five clicks he had observed in the |Xam language in addition to providing signs to indicate nasalisation of certain sounds, stops, tones and various other unique aspects of a San language which have also proved problematic to

modern orthographers. It is significant that the words of the San informants //Kabbo, /Han=kasso and Dia!wain have inspired several books of poetry based on the translations into English by Bleek and Lloyd, although the original translations are literal and difficult to unravel (Markowitz 1956; Watson 1991). It is perhaps significant that both Bleek and his daughter Dorothea used the word 'literature' in describing the folklore, natural history and beliefs which they collected, surely an indication of their respect for the people whose heritage it was.

The folklore has thrown light on the mythological aspects of San rock art and also shown the existence of a similar belief system among people widely separated by time and place. Modern ethnographic research has revealed an affinity between folktales of the modern San of the Kalahari and those of the now extinct /Xam people of the Cape. The Bleek and Lloyd collection continues to be of great value to modern researchers.

Some Recent Projects in the Collection of Oral Data

During the early 1990s, when the Ju/'hoansi people of the Nyae Nyae area of Namibia set up the Village Schools Project, a unique method was devised by means of which children were able to learn to read their own language when they entered the schools established in the seven villages. Young Ju/'oan trainee teachers went round the communities with tape recorders to interview the elders and tape the folktales they remembered. These were then used to produce beautifully illustrated reading books for the children.

Currently several communities in Namibia are in the process of collecting oral data from the elders who can recall the past way of life and the stories told to them by their parents and grandparents. This data will form an invaluable source of information for the younger generation and will enable them to develop a pride in their own heritage. In Botswana two important initiatives are the Naro Language Project and the Khwe Oral History/Testimony Programme. The Naro language is spoken by many people in Western Botswana but had not previously been written down until a project was begun under the auspices of the Kuru Development Trust to provide an orthography and translate the Bible into Naro. There are now a number of trained Naro speakers working with the missionary couple who have led the Project for some years. Several books of the Old and New Testament have been published in Naro and a monthly newsletter is provided so that Naro who are already literate in either English or Setswana can learn to read and write in their own language. As for the Khwe Oral History/Testimony Programme mentioned above, it began as an initiative of two young speakers of the //Anikhwe language, who in 1996 and 1997 began to collect information from older people in the Okavango Panhandle in northwestern Botswana. The interviews recorded on tape were translated into English and Setswana but include names of places and people in either Khwedam or ||Anikhwedam using an orthography developed by researchers from the University of Cologne (VanderPost 2003:9).

The use of modern technology has made it possible to record information orally so that future researchers of the languages will have information not available in earlier days. Speakers of the languages will then have access to the folklore of their people which in future may also be transcribed into their written languages.

Conclusion

Reading vast amounts of material while preparing the Khoe and San bibliography proved both challenging and rewarding. The indexing of these Khoe-San languages presents unusual and complex problems, which the author has attempted to delineate in this chapter.

BIBLIOGRAPHY

Andersson, L-G & Janson, T. 1997. *Languages in Botswana: language ecology in Southern Africa.* Gaborone: Longman Botswana.

Batibo, HM & Tsonope, I (eds). 2000. *The state of the Khoesan languages in Botswana.* Mogoditshane: Tassells.

Bleek, WHI & Lloyd, LC. 1968. *Specimens of Bushmen folklore.* Cape Town: Struik. (Facsimile reprint).

Crawhall, N. 2004. Personal communication.

Deacon, J & Dowson, TA (eds). 1996. *Voices from the past: /Xam Bushmen and the Bleek and Lloyd Collection.* Johannesburg: Witwatersrand University Press.

Lane, P, Hermans, J & Chadzimula, M (eds). *Proceedings from the Basarwa Research Workshop, 24-25 August, 1995.* Gaborone: The University of Botswana and the University of Tromso Collaborative Programme for San/Basarwa Research.

Markowitz, A. 1956. *Stories, myths and fables of the South African Bushmen told in their manner.* Cape Town: Central News Agency.

Monageng, S. 2001. In *Proceedings from the Basarwa Research Workshop, 24-25 August, 1995,* edited by P Lane, J Hermans and M Chadzimula. Gaborone: The University of Botswana and the University of Tromso Collaborative Programme for San/Basarwa Research.

Penduka Declaration on the Standardization of Ju and Khoe Languages. 2001. Penduka Training Centre, Windhoek, Namibia, 20-22 April 2001. (Unpublished).

Spohr, OH. 1962. *Wilhelm Heinrich Immanuel Bleek: a bio-bibliographical sketch.* Cape Town: University of Cape Town Libraries.

Treis, Y. 1998. Names of Khoisan languages and their variants, in *Research in Khoisan studies,* Volume 15. Cologne: Rudiger Koppe Verlag.

VanderPost, C. *Community natural resources of Bugakhwe and ||Anikhwe in the Okavango Panhandle in Botswana.* Shakawe: Teemacane Trust.

Watson, S. 1991. *Return of the moon: versions from the /Xam.* Cape Town: Carrefour.

Willet, S. 2003. *The Khoe and San: an annotated bibliography, Volume 2.* Gaborone: University of Botswana.

Willet, S, Monageng, S, Saugestad, S & Hermans, J. 2002. *The Khoe and San: an annotated bibliography, Volume 1.* Gaborone: University of Botswana.

Working Group of Indigenous Minorities in Southern Africa (WIMSA). 2001. The Penduka Declaration on the Standardization of Ju and Khoe Languages. Penduka Training Centre, Windhoek, Namibia, 20-22 April 2001. (Unpublished).

25 LAW A Case Study on the Services to Postgraduate Students at the University of South Africa (Unisa)

Nico Ferreira

Abstract

The services of the Unisa Law Library are analysed in order to provide the reader with information on methods of legal information retrieval that can be generally applied.

Introduction

The purpose of this contribution is to provide an insight into the retrieval of legal information by briefly examining different legal bibliographies as a case study illustrating the variety of services delivered to postgraduate students at the Unisa Library, Pretoria.

Context

The author's perspective and experience as the Law Librarian and the services offered by the Unisa Law Library form the basis of this study.

As Unisa is a distance teaching institution, the library is obliged to cater for on-campus clients, as well as those living elsewhere, both nationally and abroad. These off-campus clients may contact the library through the postal services, e-mail, faxes or telephonically.

The clients of the library can be categorised as follows:

- Undergraduates: approximately 5 000 registered law students
- Postgraduates: approximately 400 registered legal researchers on Masters and PhD levels
- Prospective postgraduates: students who have yet to finalise their registrations but who are eager to start doing research on a specific legal topic
- Unisa staff: approximately 200 law lecturers
- Postgraduate students and staff from other tertiary institutions
- Members of the SA legal fraternity
- Any community member with a legal information query or a legal research need

The Unisa Law Library is centralised on the main campus as a special collection within the Main Library. At the time of writing, the Unisa Library Service was still in the process of merging with VISTA and Technikon RSA (TSA). The contribution will therefore be based on current services and processes.

Certain factors must be considered:

- South African resources versus foreign resources. The new South African Constitution compels legal practitioners to take note of foreign legal resources from foreign legal systems.
- Access versus ownership. Based on the available information resources budget, decisions will have to be made whether a library will provide access to resources or whether it will buy the relevant resources.
- Hard copy versus electronic source. Collection developers will have to take this factor into consideration when deciding on future developments. Various options need to be examined: Will a collection become increasingly electronic-based? One should take into consideration that electronic resources are not automatically cheaper. Will the developer opt for an interim period where both hard copies as well as the electronic resources are purchased for the collection? Alternatively, will the developer retain the traditional paper resources?
- Full-text versus bibliographic sources. It seems some researchers still believe that all material is available full-text on the Internet, ready to be printed freely at the touch of a button. They are also ignorant of the fact that certain databases, for example the *Index to legal bibliographies*, are bibliographic sources that (only) function as indicators of possible resources on a particular topic.

The focus of the chapter will not be on the theoretical aspects of bibliographic resources or bibliographic services. However, one should note that the practice of any library and information service should be based on a solid theoretical foundation.

Some ethical considerations are important to the delivery of information services:

- Law librarians do not provide legal advice, but do provide information resources which may contain the solution to a legal problem.
- According to the Attorneys Act, only qualified lawyers may provide a legal service.
- Therefore law librarians do not interpret Acts or law reports;
- Client confidentiality is very important.
- Keep in mind the legal maxim that states: 'A person who acts as his own lawyer has a fool for a client.'

Different Types of Legal Bibliographies

The following types of bibliographies are available:

- In book form. For example *A bibliography of South African criminal law*, by A Rabie.
- At the end of chapters in books. For example *Annual survey of South African law*, in which each chapter discusses a different area of development in South African law development for a specific year, with a bibliography of the most important books and journal articles published in that year.
- As journal articles. For example in the 1990 *Journal for Contemporary Roman Dutch Law* there is an exploratory bibliographic article on South African law journals.

- Covering a certain time span. For example a bibliography of articles on labour law, 1989-March 1993, in the *Industrial Law Journal*, 16(5), 1993.
- As a comment on the current situation. For example on marital rape in South Africa, as published in the *South African Journal on Human Rights*, 1992.
- As a historical perspective. For example literature on the police in South Africa, as published in *Acta Juridica*, 1989.
- As a guide to legislation on a certain subject. For example, a bibliographic guide to South African insurance legislation, in *Modern Businessman*, 1988.
- As journal special issues. For example 'Elder law: a special issue', in University of Kansa *Law Review*, 1996.
- Bibliography of bibliographies. For example 'Bibliography of bibliographies on criminal violence', in *Journal of Criminal Law and Criminology*, 1984.
- As a legal overview and research bibliography. For example 'Understanding patents: a legal overview and research bibliography', in *Legal Reference Services Quarterly*, 2001.
- Bibliography to a conference. For example 'Bibliography to the Conference on the Delivery of Legal Services to Low-income Persons', in *Fordham Law Review*, 1999.
- As a research guide. For example a research guide to medical malpractice and the law, in *Legal Reference Service Quarterly*, 1997. A research guide can also be called a pathfinder and it looks like an annotated bibliography.
- Basic guide and bibliography for students and practitioners. For example 'Research of Jewish law: a basic guide and bibliography', in University of Detroit *Mercy Law Review*, 2000.
- In addition, the well-known bibliographies.

Several of the bibliographies were published some time ago – but legal researchers must look at both old and new resources, in order to ascertain how the law has developed.

Bibliographic Services

The mission of the Library is as follows:

> *The Library Service and the Unisa Community are partners in the pursuit of academic excellence. We actively support the teaching, research and community service programmes of the University by providing client-orientated, quality library and information services and training, using appropriate means. The skills, knowledge and helpfulness of the staff, the advanced information technology, and the excellent collection contribute towards maintaining outstanding academic information services.*

In the pursuit of efficient subject retrieval, the following sources and bibliographic resources have been made available.

Index to SA Legal Periodicals

This is compiled on a contractual agreement with the South African National Library by the Unisa law librarians. The records are delivered to the SA National Library and ultimately

made available on the MAGNET database by SABINET as part of the Index to South African Periodicals (ISAP).

The following is an example from ISAP as available on SABINET:

DU PLESSIS, J.
Psychiatric injury sustained by secondary victim due to watching primary victim being struck by a truck
De Rebus, no. 410, April 2002, p. 45-46.
Discusses the outcome of Road Accident Fund v Sauls 2002(2) SA 55(SCA). Looks at the relationship between the primary and secondary victims.
CASE DISCUSSIONS, LAW OF DELICT, EMOTIONAL SHOCK, CAUSATION, NEGLIGENCE, POST-TRAUMATIC STRESS, PSYCHIATRIC INJURIES, ROAD ACCIDENT FUND v SAULS

In the past it was possible to find this record by doing either an author search, article title search, journal title search, or by means of a subject search. Now it is also possible to find the article by doing an abstract search or a subject search when information pertaining to a specific case report is available (in this case *Road Accident Fund v Sauls*). This filled a need for discussion on specific court cases. The same can be said about finding information on specific South African Acts and legislation. By including additional information in both the abstract and the subject fields of the record, the relevant information will be easily retrieved.

South African Legal Abbreviations

It is well known that abbreviations abound in the field of law. It may even happen that a legal book or journal article contains abbreviations without the necessary key to these abbreviations.

A citation such as TRAN CBL is not 'wrong' or should not be regarded by librarians as 'containing insufficient information'. And naturally no librarian can be expected to know the meaning of all abbreviations – but any law librarian should know where to find the key.

The number of queries on legal abbreviations received at the Unisa law library indicated that a South African publication on abbreviations was sorely needed. This led to the publication of the following bibliographic resource:

Ferreira, Nico M & Breckon, Karen E
South African legal abbreviations
Buffalo, New York: William S Hein & Co, 1999
ISBN 1575885263

Thesaurus of South African Legal Terms

Compiled in cooperation with other law librarians in South Africa, this thesaurus indicates that indexing is a subjective exercise and that a need exists for a publication to assist with finding the correct search terms when searching for legal journal articles in ISAP (on SABINET).

The following is a page from the draft thesaurus:

Judiciary,	
single	use Unified judiciary
Juries	also Jury system
Jurisdiction	more specific Jurisdiction of the courts more specific Territorial jurisdiction
Jury system	also Juries
Justice,	
administration	use Administration of justice
Justus error	also Error in law also Iustus error
Kennisleer	see Doctrine of notice
Labour law	
(very broad)	also Employment law (very broad)
Labour unions	also Trade unions

Note the inclusion of several languages and the inclusion of different references, such as *also*, *use*, *more specific* and *see* – all to help the researcher find the correct term in order to locate all relevant journal articles.

The examples above illustrate Unisa Law Library staff's dedication to making information more readily accessible.

Subject Librarians

At the Unisa Library a subject librarian system is in place whereby at least one subject librarian is available for each teaching department or group of departments. There are two full-time subject librarians for the Law Library, as well as three assistants who staff the information desk.

Services to Clients

Compiling Reading Lists

A reading list may be needed for several reasons:

- for new Masters and PhD students at the start of their research
- for promoters who are in communication with the researchers
- for researchers working on a new book or journal article and who need the newest information on the specific subject
- for lecturers compiling new study guides or updating them

The starting point for compiling a reading list is the library catalogue, which can be found at http://www.unisa.ac.za. The complete library collection can be found here. To find references to legal journals, use (as only one example) the *Index to legal journals and books,* published by WH Wilson, and for foreign journals, the index to *Foreign legal*

journals, produced by the American Association of Law Libraries. South African journal articles are located on the already mentioned ISAP.

The library subscribes to an extensive collection of databases which can be browsed electronically or on hard copy.

Answering Queries

No librarian could possibly have all the answers, but should be able to find sources that provide answers to a question or at least know in which direction to steer a client. Being able to answer queries is a combination of experience, knowledge about the subject and which bibliographic source to use. It is often necessary to remind the client that books do contain indexes and that the index itself is a way to find specific information.

Finding Bibliographic Information

This can be a very testing exercise, especially when a postgraduate student or researcher is completing a thesis and is trying to find elusive bibliographic information for the footnotes or the bibliography.

Training

The training of clients in the use of bibliographic information resources is a major component of any subject librarian's daily routine. This training may be on an informal one-on-one basis or take the form of more formal training sessions in a modern electronic learning centre.

Collection Development

Unisa library is fortunate that the subsidy it receives from the national government is based on the total number of registered students. At the time of writing, there are about 125 000 students, with the result that the budget available for collection development is quite substantial. However, it is impossible for one law library to collect every single law publication. This activity is a very important task in the daily routine of any subject librarian. We often compare reading lists compiled from foreign legal resources with the library collection to see which resources are not in our own stock. The importance of the role of legal bibliographies in law collection development can never be overemphasised.

Conclusion

The following quote by SB Pagel (1989:9) refers to legal bibliographies and law librarians (although it rings true for all subjects and all librarians):

> When all is said and done, when the state of legal bibliography past, present, and future has been explored, what role should law librarians play in producing bibliographies? They have a duty to become involved with the needs of their users to

a major degree through compilation and publication of these tools. Legal information professionals have special skills, and legal researchers have special needs, which require that librarians participate in creating guides and bibliographies. Members of the profession have a duty to use their skills to the benefit of their patrons, their institutions, and themselves. The legal bibliography is alive and well – it remains a vibrant, healthy genre which will continue to grow and evolve, and sometimes even appear in a different format, so long as there are librarians interested in serving the research needs of the legal community.

BIBLIOGRAPHY

Pagel, SB. 1989. The legal bibliography. *Legal Reference Services Quarterly*, 9(1):9.

26 LITERATURE FOR CHILDREN

Thomas van der Walt and Karin McGuirk

Abstract

After indicating the importance of indexes for children's books, the authors focus on South African publishers of children's literature and their attitude towards the inclusion of indexes. They summarise their findings and conclude that although the majority had a positive attitude towards indexes, financial and other constraints often preclude their inclusion.

Introduction

In the school environment much attention has been paid lately to computer, and especially Internet-based, information searches. It is often forgotten that a large and very important body of information is still available in printed format. Certain means are necessary to guide the user in finding this information and using it properly. In this case it is the young user still finding his or her place in a world where everything revolves around the timely location of relevant information. This includes a focus on the skills needed to be able to digest the found information, as well as the application of the processed information to fulfil specific needs. The two most obvious key means are the library catalogue, specifically the subject catalogue, and book indexes. Book indexes are supposed to serve as a key to the exact place where detailed information on a specific subject in a book (printed information source) can be found. A book index works in a way similar to the subject catalogue of a library, indicating where information on a specific subject can be found in the library (Gordon 1976:294). Yet the crucial role that indexes in information books for children can play is often underestimated and many such books are published without an index. It is often said that there is no need for indexes in simpler books for children.

The Role of Indexes in Books for Children and Young People

The most successful person, according to Marland (1992:12), is someone who has the ability to turn a need into a question. To answer the question, such a person will be able to search for relevant information. The search process includes the selecting, organising, evaluating and presenting of information as an answer to the need-turned-question. To be able to do this, such a person needs to possess certain basic information skills.

Such skills are essential during the first three to four years of schooling. In this period children become especially familiar with the skills of judgment and differentiation. Children are immensely curious about everything around them, and during this period their

experience of the world around them stimulates the need to know more. All of this comes down to the analysing, organising and reorganising of information.

These skills form the cornerstone of outcomes-based education (OBE). Learners are expected to be able to do autonomous as well as group research. They must be able to find information, and process and repackage it. Learners must not only be taught how to read; they must also be taught to read with a view to becoming independent learners. As independent learners they will for the rest of their lives be able to select, analyse and synthesise information in an independent manner (Marland 1978:68). Acquiring information skills therefore plays an important role within the OBE curriculum. A survey undertaken in June 2004 among media teachers in Pretoria (Van der Walt 2004) indicated that all media teachers felt book indexes play an important role in the acquisition of information skills. They also indicated that book indexes are employed to show learners how to identify key terms when searching for information and from there, how to locate relevant information. The index remains a basic resource for teaching learners how to do simple research. It encourages autonomous learning as well as an orderly and systematic approach to learning and information retrieval.

The above survey strongly indicates that media teachers understand the value of the index, and that it is taught and implemented as an information skill. It was specifically mentioned as one of the basic information skills required. All the primary and secondary school media teachers participating in the survey also indicated that learners were not only aware of what an index is, but also understood that an index would help them find information more easily and more quickly. Most of the children were also aware of the difference between an index and a table of contents. The validity of the teachers' assumptions has not been scientifically tested, and one teacher stated: '[R]egular library users know what an index is and how to use it, others simply want a book on a specific topic and must be shown how to get their information from different sources. They actually use keywords to search electronically, but then the system does all the work.'

Characteristics and Standards for Indexes for Children

Because indexes are so crucial for children, a set of basic standards with which a good index should comply, is needed. According to Marland (1978:68) an index has three functions: to indicate the scope, the level and the detail of the book. In other words, it indicates in the first place the appropriateness of the specific source to the purpose for which the user needs it. Secondly, it allows for a 'disorganising' of the author's particular organising of the information. This allows users to create their own reorganising of the relevant information. In the third place, it serves as a means of retrieval. These functions, though, can only be used to their full potential if the index has been properly constructed and complies with particular characteristics. Indexes in information books for children are constructed and offered in different ways. This makes its use difficult for learners and is sometimes found too confusing to use. However, we also need to avoid becoming prescriptive because at present there is no consensus among indexers, authors or media teachers on specific issues.

An index in a children's book, as with any other index, should be able to stand as its own entity and not simply appear as an afterthought at the end of the book. When constructing an index for a children's book, an indexer must constantly keep in mind the child as the user. Some of the questions to consider are what knowledge the expected user will have of the book, which indexing terms should be familiar to the user and how the user will search for information. An index for a children's book should not be a simplified version of an index for grown-ups. The introductory notes and information given must be informal and user-friendly, and the terminology used needs to be suitable. Appropriate headings must stand out in prominent print and the typography must not be smaller than the main text, so that children do not see the headings as less important than the main text.

One of the issues with indexing for children about which there is no consensus is which method of arrangement is better: the word-by-word or the letter-by-letter method. Most media teachers approached for the purpose of this study felt the word-for-word method was best. But others preferred the letter-by-letter method since children tend to search strictly according to the alphabet and may find it confusing to search according to words. However supporters of the word-by-word method felt this method made it easier for learners because 'the entire word is mentioned'! Yet what is probably more important than choosing between the word-by-word and letter-by-letter methods is to be consistent. Sometimes even older children give up if there is no clarity and consistency in the presentation of subjects in an index.

Concerning the format of the entries, the indexer needs to make a calculated guess about which method most users will probably use. For example:

```
seals, hunting of

or
seals
    hunting of

or
hunting
    of seals

Blood River, Battle of

or
Battle of Blood River
```

Local media teachers consulted expressed no definite preference for one or the other of the two formats.

A constantly debated issue about indexes for children's books concerns page references. For example: '4-6' or '4,5,6'? Media teachers interviewed had no specific preference (Van der Walt 2004). However Mathews and Bakewell (1997:193) mention that an indication from one particular page to another particular page (e.g. 4-6) can confuse a younger child, who may even ask 'What does 4 minus 6 mean?' So it may be better to list all the pages: 4, 5, 6. But of course a page indication like 4-6 clearly indicates uninterrupted coverage of the

specific subject from page 4 to page 6 – which means there will be more information on the subject on these pages than there would be in 'passing references', that is page references where the subject may only be referred to in passing on the page mentioned. The page reference '4, 5, 6' therefore does not distinguish it from 'passing references'. It is agreed that 'passing references' must be avoided, and even more so here than in books for adults. The general feeling is that passing references in children's books may do more harm than good. It is, or could be, already problematic for children to find the right indexing term. From there they need to search for the relevant information on the pages mentioned, a process that can go slowly. After a search through a long list of references for a subject, a child may find some pages do not lead to any useful information. This will only discourage any future use of the index.

In general, there should not be too many page references for single entries in indexes for children. Media teachers who participated in the survey supported this view. Indexes aimed at older children can be more comprehensive, but references to subjects only mentioned in passing should still be avoided (Van der Walt 2004). It goes without saying that abbreviations such as '4ff' should not be used.

One way of avoiding long strings of page references under a specific term is to use subheadings. For example:

Mandela, Nelson
    birthplace
    president
    prisoner on Robben Island

Local media teachers think subheadings simplify the research process and help learners find out where to access specific information immediately (Van der Walt 2004). Subheadings have one major drawback that can also lead to confusion: they disrupt the general alphabetical order of terms with a new alphabetical order under a main heading. The indexer should weigh up the benefits and drawbacks. The use of too many and unnecessary subheadings can be confusing, but then too many long strings of page references under one heading may be equally confusing and discouraging.

There is as little consensus about cross-referencing in indexes for children. According to a British study (Mathews & Bakewell 1997:193), most children usually find cross-references difficult to understand. The opinions of local media teachers were divided on the use of cross-references, especially regarding works for junior learners. It was felt cross-references must be made very clear, if they are used. Perhaps a better solution would be to use multiple entries. Children do not yet have enough linguistic knowledge to be able to search under synonyms if they find that the term they originally searched under is not in the index.

To make it easier to locate specific indexing terms, it is better to use a space to distinguish each first indexing term starting with a new letter of the alphabet from the terms starting with the previous letter. Another possible way to distinguish between indexing terms is to clearly indicate each alphabet letter (lower and upper case) at the beginning of the terms starting with the specific letter, preferably in bold. Even though some of the local librarians

consider the open spaces more important than the print of the letter itself, most others think both methods should be used simultaneously (Van der Walt 2004).

The general view in the literature on indexing illustrations is that it is important, but should be distinguished clearly from text references. This can be done using typography – for example printing in bold or italics (Gordon 1976:298). It is interesting to note that local media teachers do not have strong views on the matter. Most indicated that while it can be done, they do not think it crucial. Responses that it should only be done if the illustrations are important (yet without specifying what is and what is not important about an illustration), or that it is only useful to index illustrations which users will want to search for, are not really meaningful (Van der Walt 2004). Obviously it is unnecessary to index illustrations that appear on the same page as the text covering the same subject.

The Table of Contents as Distinct from the Index

A glance through a book's table of contents should give one an idea of the contents and main themes. The longer and more detailed a table of contents, the less helpful to the reader. It is the purpose of the index to represent the book in detail. This representation is not in the order that it appears in the text, but in an alphabetical order that makes it easier to locate specific information. According to Gordon (1976:294) many publishers feel a detailed table of contents offers sufficient search possibilities for children and an index is therefore unnecessary. But a table of contents is arranged thematically and arbitrarily – so clearly this is not the case. For example, a child searching for information on polar bears may think of searching under a term such as 'Mammals' in the table of contents. But there will be problems if the table of contents happens to be arranged geographically, since this child is not likely to think of searching under 'North Pole'. An index should not offer any problems in a similar search process.

The local media teachers agreed that the table of contents does not have enough scope for information searches. The two main reasons why all media teachers find the inclusion of indexes in children's books necessary is that a table of contents is not alphabetical and does not represent all the information contained in the source (Van der Walt 2004).

Who Should Construct the Index?

It is clear from the literature on indexing for children that it is a subject of ongoing discussion. In practice many authors construct their own indexes, although it is generally felt the task should be entrusted to professional indexers.

However there are definite benefits to the author's taking responsibility for the index, mainly because the author is familiar with the content of the book, as well as with the subject material in general. The author will also be aware of the needs of the book's particular target group. Usually authors of information books for children are well aware how having an index will improve the book's marketability. A proper index is conducive to good sales (Mulvany 1994:28).

But it is also true that indexing is an art and a skill, and that generally authors are not properly conversant with the indexing process. An author who has worked long and

intensely to finish a book (writing and rewriting, editing and proofreading) is often left exhausted at the end, with little energy to compile a proper index. The result is a substandard product. Mulvany (1994:28) emphasises how the construction or 'writing' of an index, as a writing style, differs in a marked way from a creative or 'expository' writing style. An author therefore needs to adapt to the required style and an author who is new to indexing obviously needs to get to know the main characteristics of an index and the process of indexing. Seal (1984:113) indicates that 60 per cent of authors write no more than one book – so for these authors there is little point in making the effort to learn the skill of indexing. Moreover, there is no guarantee that an author who knows the language of his or her subject well will find this language suitable or usable for an index. The author may have used specific terms and concepts in the text when writing about the subject, while other terms and concepts may offer better access to the specific subject material. If an author is also responsible for the index it is important to dissociate himself or herself from the book. This is especially difficult if this author has been working on and 'living with' that book for some time. However, in spite of all the obvious problems associated with indexes done by authors, Mulvany (1994:28) still feels authors who do wish to do the indexing themselves should be encouraged to do so, provided they are aware of the challenges.

The most important benefit of indexing done by professional indexers is that they have the necessary experience. This enables them to plan their time so that they finish the index in time. They tend to be more objective about the book and come to the text with a fresh outlook on the subject material. Indexers should therefore be able to create an index that is a 'special blending of the language of the text with the language of the audience'. This, according to Mulvany (1994:28), is one of the most important characteristics of a good index.

Irrespective of who is responsible for the index, it should be compiled with great care. Jean Kirk (1988:78) writes the following in an article on users' approaches to indexes: 'It becomes clear that the user, and the indexer, work on preconceptions and look under the heading that seems logical to search or index by the defined, accepted rules, ... [T]hese preconceptions develop from experience and training so are unavoidable. To what extent, then, is the indexer to take into account the likely approach of the user, in trying to make the contents of books ... retrievable by listing names, places and subjects in a predictable order, with an indication of their physical position within the source? Is it possible for the user to understand the indexer's approach?'

The indexer of children's books has no choice. He or she must take into consideration the possible approach of the user. Child users cannot be expected to be able to consider the indexer's approach when they search for entries in the index.

If an indexer is responsible for the index and not the author, then cooperation with the author is still essential. South African authors are often responsible for compiling indexes (Breytenbach 2004) or sometimes responsible for finding an indexer to do the compilation. The quality of the index is a reflection on the author and publisher. It is therefore imperative to find the right person – someone with whom cooperation and communication is possible and preferably someone who has previous experience of the particular type of book. Of course the author should be available for consultation and can serve as a valuable technical source for the indexer. However, the author should not get too involved. According to

Mulvany (1994:29), authors often give indexers long lists of terms that they want included in the index – but this is often unnecessary and counterproductive. It is more important to discuss the specific terminology and for the author to point out any specific points that need to be accentuated. Mulvany (1994:29) also mentions that authors should let publishers know as early as possible exactly how involved they will be in the indexing process. One activity for which the author's involvement is essential is the 'review' of the index; this way the author can determine whether certain subject material is handled as planned or represented in the text. The author should be able to assess whether the index succeeds in giving access to the content of the book. A random piece of text can be selected as a test to visualise how readers may search for the information contained in it. The aim is to determine which indexing terms they may use. Then consult the index and see whether these terms have been included. The index needs to be checked or reviewed as a complete unit, for example to see whether indexing terms are used in a consistent manner, that cross-references are logical and necessary, et cetera.

This review process comes with its own problems, especially since indexing is not a linear process, but rather a developing or evolving one. The index changes as the indexer works on it and, when an author wants to check on an interim basis whether it complies with his or her requirements, this may interfere with the indexer's thinking pattern unnecessarily.

Current Practice in South Africa

A survey among South African publishers indicates that most are aware of the importance of indexes in children's books, although some do have reservations about just how important. Miemie du Plessis (2004) of Lapa Publishers stresses the importance of indexing when she points out that it is the easiest way to find information and that already at a young age children should know the index is the first point of search. She frankly admits that it should be publishers' policy to include indexes in all the information books they publish (local publications as well as translations of overseas work), but that this does not always happen in practice. Ney-Her (2004) of Awareness Publishing feels children's book publishers have a duty to make sure indexes are included in their publications, for the same reason as given by Du Plessis that children can learn from a young age to use an index. Awareness Publishers always include indexes in their information books for children (Ney-Her 2004). Oxford University Press publish few information books for children, but when they do an index is included if appropriate (Spoor 2004). They are very much aware of the value of an index 'because of its value as a vital tool for developing information retrieval skills, not just for school work but for life' (Spoor 2004). Similar to them are Maskew Miller Longman, for whom the inclusion of an index is determined by the nature of the book, the subject or learning area of the book and the grade it is aimed at (Van Zyl 2004). The nature of the book is the determining factor for the publisher and in their case it is always books intended for South African schools. The feeling is it is less useful to children under Grade 6 level. 'It should, however, not be intimidating at that level and children should be taught how to use it' (Van Zyl 2004). Macmillan Publishers always include indexes irrespective of the feeling that children do not really use them (Weber 2004). Linda Weber (2004) of Macmillan Publishers believes indexes are underused because learners do not get

proper training on their use and value. 'Given the low level of literacy generally and the importance of developing information literacy in particular, I think it's crucial that learners be exposed to and even drilled in the use of all tools that facilitate their ability to retrieve and access information.' This remark from a publisher responsible for many information books indicates that the rose-tinted picture sketched by local media teachers is definitely not applicable in all instances.

The only exception is Human and Rousseau Publishers. These publishers do not publish many information books. Their children's book editor believes an index is not important because the text for children's books is supposed to be a lot more ordered and systematised than for adults (Lategan 2004). If a children's book does have an index the feeling is it should be simple but broad enough and easy for children to use. It also depends on the book's content (Lategan 2004). Yet in practice, this particular publisher is not consistent in the inclusion of indexes and the editor does not see its value. Recently Human and Rousseau brought out two information books, one a book on South African birds including an index. The second book, on cricket, only has a detailed table of contents, which according to the editor has worked out quite well (Lategan 2004). Lategan (2004) clearly states that in her opinion a simple list of terms is more important than an index, and also mentions she prefers an annotated table of contents.

It seems most publishers of information books for children are aware of the importance of a professional indexer having responsibility for the index, but that several concerns impact on its execution. Oxford University Press always use a professional indexer for more complex texts. Generally, however, it is the author, in consultation with the editor, who will construct the index (Spoor 2004). Awareness Publishing will use a professional indexer with pleasure – 'if you can find one'. In practice, it is often the publisher's copy editor who does the indexing (Ney-Her 2004). Macmillan Publishers feel the nature of the information books published (mostly primary school textbooks) makes the services of a professional indexer unnecessary. The publisher asks the author to supply a list of the most important terms and concepts along with the manuscript, and this is passed on to the typesetter. 'The typesetter builds the index using the indexing feature in software like Quark' (Weber 2004). She still sees the value of a professional indexer because she also indicates that the publisher will most definitely use the services of such a person when it comes to publishing information books at tertiary school level, or books of a highly specialised nature (Weber 2004). Even though Human and Rousseau's children's book editor indicates on the one hand that indexes are not really necessary in such books, she does feel on the other hand that professional indexers should be responsible when an index is needed (Lategan 2004). The nature of the book and the index, for Maskew Miller Longman, determine who will construct the index (Van Zyl 2004):

> The general rule is that a professional indexer will be asked to do the indexing, especially the more complex indexes. However, it also depends a lot on the subject matter or learning area. For Geography and History books for example, a professional indexer will always be appointed but if it is a very basic index for a language book, the editor might be asked to do it. If it is a very basic one-page index for a children's book, the editor might be asked to do it. If it is a fairly inexperienced editor, we will

not ask him or her to do the index. In some cases the authors might even supply it themselves.

At Lapa Publishers, the editor, or the author, constructs the indexes – to cut costs (Du Plessis 2004).

When professional indexers are used, the importance of cooperation between indexer and author is stressed. Maskew Miller Longman do the following: '[The editor or in-house project manager] briefs the indexer on what to do and what is expected. The indexer is told how many pages the index should make, e.g. 4 pages in the printed book, etc. It is also important that the indexers know exactly who the market is, e.g. whether it is for G6 learners or G10 learners' (Van Zyl 2004). Weber (2004) of Macmillan mentions that the 'indexer must work to a brief which takes on board key factors such as the age/grade of the learners, the predominant teaching and learning area of the title, the home or first additional language of the intended audience, etc.' Awareness Publishing limit cooperation to 'the normal checking of the other' (Ney-Her 2004). Human and Rousseau may not have any real experience in this field, but the children and youth book editor feels there must definitely be cooperation. In such a case the publisher would prescribe the depth of the index, for example indicate the purpose of the publication. The depth and aim of the index (or annotated table of contents) will also be determined by the text layout (Lategan 2004). Awareness Publishing and Lapa Publishers feel instructions must be clear, brief and to the point (Du Plessis 2004; Ney-Her 2004).

The different approaches of the various South African publishers to the importance of indexing are also reflected in their views on what knowledge an indexer should possess. As expected, in view of the value they attach to indexes, Maskew Miller Longman attach particular importance to an indexer's existing knowledge. Van Zyl feels this is especially important for books dealing with subject areas like economics, geography and history, and other technical subject fields. They use professional indexers mainly, who therefore know how to construct an index, and this is the most important thing for the publisher. 'However, as far as the age group is concerned, they definitely need to know who the book is for, intended market, etc. An index for G6 learners that is too complicated or comprehensive will confuse learners instead of helping them' (Van Zyl 2004). Lategan (2004) of Human and Rousseau finds good indexers are scarce and if an indexer does not have the necessary knowledge on the specific subject, then he or she will work closely with the author and the editor. It is strange that the editor clearly indicates that knowledge of the specific target group of the book is not necessary, and that the author and publisher will give guidance. Lapa Publishers feel the indexer must definitely be very familiar with the book's subject and target group (Du Plessis 2004). Ney-Her (2004) of Awareness Publishing and Weber (2004) of Macmillan both feel that this knowledge is not necessary. Weber (2004) does indicate that her publisher may appoint an editor to check the index to make sure that it is relevant and representative of the curriculum. The Oxford University Press editor mentions that a professional indexer must be able to work on any topic, but that the publisher should use a specialist when it comes to more specialised subjects (Spoor 2004). It should be ascertained whether an index is usable by children, according to Weber (2004), in the following sense: '[T]he index should reflect subject-specific, content-related information and not arbitrary

references. For example, an index in a book on the history of aviation in South Africa should not include a reference to some public figure who happened to fly in a particular aircraft and who is mentioned only in passing or anecdotally in the text.'

There is also no agreement whether there is a difference between indexing for children and indexing for adults. Human and Rousseau and Lapa Publishers are very specific that there is a difference (Du Plessis 2004; Lategan 2004). According to Lategan the difference lies in the simpler approach (fewer entries) of children's books, and that hit words must be indicated more clearly. Du Plessis refers to the choice of words and the degree of difficulty or depth of the index. This contrasts with the opinions expressed by Awareness Publishing (Ney-Her 2004), Macmillan (Weber 2004), Maskew Miller Longman (Van Zyl 2004) and Oxford University Press (Spoor 2004). Weber (2004) says: '[A]ll indexes should be sensible. Books for adults on specialist themes would have more complex, tiered indexes, but that would be dictated by the subject matter of the book, rather than the age of the reader.' In Spoor's (2004) opinion information retrieval skills must be built on a sound base, therefore with no different approaches to the construction of indexes for children. But she does add: '[A] children's book may well approach it from a more concrete rather than conceptual stand point – one must address the age and needs of the learner.' Van Zyl (2004) bases her view that there should not be any differences on the fact that children should be made familiar with the concept of an index gradually: 'I think the important thing for children is to learn what the purpose of an index is and how to use it.'

Conclusion

It is generally agreed that indexes add value to information books for children, and that it is essential for children to learn how to use indexes. As Marland (1978:69) said, 'handling an index is a central part of reading for learning'. As with any index in a book for adults, the index in a children's book serves as a 'map' or key to the subject material in the book. It plays a significant role in the acquisition of a body of information skills that each child needs to own. Using an index is not an easy skill to acquire, which makes it so important for indexers to put extra effort into its construction. It is a positive sign that most South African publishers are aware of this. Unfortunately there are factors, mainly financial, prohibiting these positive attitudes from always being applied in practice.

BIBLIOGRAPHY

Bakewell, KGB & Williams, PL. 2000. *Indexing children's books.* Sheffield, UK: Society of Indexers.

Booth, PF. 2001. *Indexing: the manual of good practice.* München: Saur:298-316.

Breytenbach, K. 2004. Personal communication. 2 September.

Dixon, Y. 1996. Indexing for children. *The Indexer,* 20(1):8-10.

Du Plessis, M. 2004. Correspondence with TB van der Walt. 1 June.

Gordon, C. 1976. The Indexing of children's books. *School Librarian,* 24(4):294-299.

Lategan, A. 2004. Correspondence with TB van der Walt. 1 June.

Marland, M. 1978. Reading to learn and using the index. *The Indexer,* 11(2):68-69.

Marland, M. 1992. The curriculum, the library and the librarian. *Education Libraries Journal*, 35(2):19-27.
Mathews, PL & Bakewell, KGB. 1997. Indexes to children's information books. *The Indexer*, 20(4):193-194.
Muller, A. 2004. Correspondence with TB van der Walt. 24 June.
Mulvany, NC. 1994. The author and the index. *The Indexer*, 19(1):28-30.
Ney-Her, M. 2004. Correspondence with TB van der Walt. 2 June.
Seal, A. 1984. Indexes from a user's viewpoint. *The Indexer*, 14(2):111-113.
Spoor, J. 2004. Correspondence with TB van der Walt. 2 June.
Stirk, J. 1988. User approaches to indexes. *The Indexer*, 16(2):75-78.
Van der Walt, TB. 2004. Unpublished report on views of Pretoria media teachers on indexing for children.
Van Zyl, A. 2004. Correspondence with TB van der Walt. 3 June.
Weber, L. 2004. Correspondence with TB van der Walt. 9 June.

27 MUSEUM OBJECTS

Malene Schulze

Abstract

The museum has a unique role as a repository for three-dimensional objects gathered from both natural and human-made environments. The very act of collection removes objects from their context and, although many of them are inherently interesting or aesthetically appealing, it is the close inter-relationship with their environment which increases their usefulness, and their enjoyment by the museum public. Preserving the non-intrinsic information about an object, such as where it comes from, who found or used it, and what it was used for, is the responsibility of the holding institution. For this reason the documentation of museum collections is vital (Stone 1992:213). It is the indexer's task to unlock this wealth of documented knowledge for potential users.

Introduction

The indexing of museum objects has many similarities to the indexing of written material, but there are many differences as well, and these present some interesting and challenging problems. The main one is that while most indexing tasks start with a block of text, museum cataloguing and indexing starts with a box of objects. These may have words on them but usually the words have to be supplied (Will 1993:157). The best way to explain what is meant by indexing in relation to museum objects is to quote Smither (1990:11) describing documentation processes at the Imperial War Museum in the United Kingdom: '[W]e catalogue the collections and then index the catalogues.'

Museum cataloguers have a lot more discretion to determine the form and content of a record than book cataloguers do. All museum cataloguing is original cataloguing and entails considerable research and analysis in order to generate a single record (Stanley 1989:130). Information about what an object is called, what it is for, and when and by whom it was made, has all to come from outside sources – either documentation accompanying the object, separate reference works, or the cataloguer's own knowledge. The records created for museum objects then require a whole network of links to associated entities such as people, organisations, places, events, cultures, materials, images and documents. These links are not static, because an object's history is not static. New information may be discovered, but even without this an object will be exhibited, stored, moved, lent, conserved, photographed and studied. All these entities and processes have to be recorded in such a way that they can be found and also used as retrieval keys to the other entities to which they are linked (Will 1997:36).

What are the contents of the museum catalogue record? Notwithstanding the fact that museums are diverse institutions housing objects from various disciplines, the following information groups are always to be found in catalogue records: *physical appearance* of the object supplemented with *non-intrinsic detail* such as geographical area where it was collected and many historical associations such as manufacturers, users and dates. Finally, the *management* of the objects such as acquisition source and locality is also recorded. (See Figure 1 in the addendum to this chapter for a list of catalogue information categories published by ICOM in the *Handbook of standards: documenting African collections* 1996.)

These days most museums use computers to record the information about their collections and this chapter is written with that approach in mind. However, the principles of information science remain the same for both manual systems and computer applications. Converting manual records to digital format is a time-consuming task which is usually implemented in phases; therefore many museums may for some time live with both systems simultaneously.

Indexing Objects rather than Text

In theory, the process of indexing museum objects should be straightforward: a curator should write a description of the object in whatever terms seem natural and appropriate, and an indexer can then work on this textual description as on any other text, providing standardised access points under controlled terms at a level of specificity appropriate to the collection. Systematic retrieval is thus provided for and when an item has been identified as being of interest, the curator's description of it is available. In practice, the distinction between description and access points is not so easy to maintain. Will (1993:157) elaborates as follows:

> To take an example from the Science Museum, we have an object with the following semi-free-text description:
> 1986-1502: Experimental bed for continuous weight measurement during psychic experiments, used by Prof. J B Hasted in his research, English, c.1980.
> There are several categories of index term by which this object might be sought: what it is, the disciplines or areas of study to which it relates, and where, when and by whom it was built and used. Appropriate indexing terms might be:
>
Object name index:	*Beds*
	Experimental apparatus
	Weighing equipment
Subject and discipline index:	*Levitation*
	Parapsychology
	Psychic research
	Weight measurement
Person/organisation name index:	*Hasted, Professor J B*
Place index:	*England*
Date index:	*1980*
>
> The distinction between description and access points is not so easy to maintain. Curators commonly do write free-text descriptions of objects, particularly when preparing scholarly or evaluative catalogues, but it is often desirable for them to provide structured descriptions too, with information broken up into fields and groups.

This: (1) allows data to be selected, sorted and presented in different formats; (2) acts as a reminder to the curator of data that ought to be included; and (3) makes it easier to scan a list of records looking for or comparing specific items of information.

Recording of data in a structured form blurs the distinction between description and access points, and there is a temptation to make a single set of data fields perform both functions. They can often do so, and then there may be economies in the amount of data stored. The lack of distinction has, however, led to enormous confusion, with curators complaining that indexers and other information staff were trying to force them to use terms that are anachronistic or that do not reflect scholarly discriminations, while indexers complain that curators are not prepared to accept the standardization of terminology which is needed to make an indexing system reasonably efficient. For most museums, the curator, the cataloguer and the indexer are the same person, who has to reconcile the conflicting points of view.

Data Standards

Much work has been done internationally over the past 20 years to develop agreed structures for museum data and terminology. Data standards not only promote the consistent recording of information but are also fundamental to the efficient retrieval of information online. They promote data sharing, improve the management of content and reduce the redundancy of effort (*Cataloguing cultural objects* 2004: Introduction).

Working parties of the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM) have developed the *CIDOC guidelines for museum object information: the CIDOC information categories*. These guidelines include information categories for collections management as well as object description. They make recommendations on syntax or format, and controlled vocabularies. The *CIDOC guidelines* are recommended both as the basis for an international standard and for new national standards, if there are no current standards in a country, and as an intermediary when comparing (mapping between) existing museum information standards (CHIN website 2002).

In the United Kingdom the lead has been taken by the Museum Documentation Association (MDA). The MDA's work over the last 20 years has made museum staff much more aware of the potential of their collections as a source of information and of the need to organise that information so that it can be retrieved effectively (Will 1993:158).

On the African continent, efforts to standardise museum data culminated in the publication of a handbook of standards in 1996, being described as 'the result of a four-year effort of the professionals of six African museums and the ICOM International Committee for Documentation (CIDOC)' (*Manuel de normes* 1996:3).

In South Africa, there is at present no national drive to establish museum documentation standards – regarding neither data structures nor terminology. In the 1980s some pioneering work was done by the Documentation Group of the South African Museums Association (SAMADOC) who distributed a 'proposed' museum documentation data standard in 1987. This is a complete listing of all categories which might be found in any kind of museum documentation and endeavours to assign order to the elements or categories and

their constituent parts. It also specifies the coding to be used for each element (*Museum documentation standard* 1987:1). Unfortunately there is no published information available on the usage of this data standard in South African museums. No final standard has been published since 1987 (Immelman 2004).

On a regional basis, the (now defunct) Transvaal Provincial Museum Service succeeded in standardising the museum catalogue records of their affiliated museums from the 1970s to the 1990s by developing and distributing cataloguing manuals. Attention was also given to the naming of objects so that a central catalogue with standardised terminology could be established at the Museum Service Head Office. Unfortunately this project was discontinued in 1994 because of financial constraints and changes in the political dispensation in South Africa.

Why are museums lagging behind on standardisation of records? Published standards are available for activities with a tradition of cooperation such as library cataloguing, which is based largely on formal characteristics of items. In contrast, museum cataloguing incorporates substantial scholarly interpretation of unique items and it may be argued that the diversity of museum collections and the unique nature of many museum objects present considerable problems for standardisation (Wentz 1995:201).

Although a data structure is the logical first step in the development of standards, a structure alone will achieve neither a high rate of descriptive consistency on the part of cataloguers, nor a high rate of retrieval on the part of end users. The choice of words (data values) and the selection, organisation and formatting of those words (data content) are two other types of standards that must be used in conjunction with an agreed-upon data structure (*Cataloguing cultural objects* 2004: Introduction).

Object Names

Fields and Structure

The MDA standard includes fields called NAME, SIMPLE NAME, FULL NAME, OTHER NAME, CLASSIFIED NAME, CATEGORY, CLASSIFIED CATEGORY, FULL CATEGORY, SIMPLE CATEGORY, DESCRIPTOR and CONCEPT, among others. Many of these fields can have further sub-structure. Although the Standard gives definitions, the distinctions are difficult to convey to non-specialists.

Although the MDA data standard allows an object to have several occurrences of SIMPLE NAME, many museums have imposed a limitation of one SIMPLE NAME per object. This allows lists to be produced in which each item appears only once and a single name can even be used for the physical arrangement of objects in storage, shelving all the teapots together, for example. ... The FULL NAME field is sometimes used to contain terms that qualify or amplify the SIMPLE NAME (Will 1993:158).

For example:

Simple name:	Submachine gun
Full name:	Auto-ordnance, Thompson Submachine Gun M1, .45
Simple name:	Medal, British
Full name:	British war medal (1914-1920)

Will (1993:158) states that such distinctions are more trouble than they are worth:

> In particular, dividing index terms into many different fields means either that each field has to be specified separately when constructing a search statement, or else a combined index has to be prepared, which contradicts the argument for separating them in the first place. Most thesauri used for indexing bibliographic material have no problem in accommodating different types of term, including names of objects, disciplines and activities; these are not hierarchically related but it is convenient to be able to make associative relationships ("related term" or "see also" references) between them.

When assigning an object name to natural science specimens, the term may be repeated to record both the varying levels of classification required (genus, species, etc.) as well as the common or simple name describing the specimen (skull, bone, etc.) (*CIDOC guidelines* 1995).

It is possible to repeat multiple object names for an object, for example 'mug' and 'souvenir', to indicate different characteristics of the same object (*CIDOC guidelines* 1995).

Language

When indexing the heritage of specific cultural groups the recording of vernacular object names or vernacular terminology for concepts is important.

Several language policy decisions have to be made for museum databases in countries with more than one official language. For instance, in bilingual countries descriptive cataloguing as well as indexing may be produced in both national languages. Alternatively, the descriptive cataloguing might be in one official language while the indexes are produced in more than one language.

The multicultural/multilingual status of a country is often reflected in its museum records. The National Cultural History Museum's acquisitions registers and catalogue cards bear witness to this. Its very first object recording in 1892 (called the entry register) was in Dutch, made by the staff of the Staatsmuseum which is the predecessor of the National Cultural History Museum. In 1903, after the Anglo-Boer War, when the museum's name was changed to Transvaal Museum, the recording language became English. The Museum's first history section register (when it was still part of the Transvaal Museum) was started in English but after the first 3 372 entries, the new registrar switched it to Afrikaans. For some reason the last 20 entries are once again in English (Grobler 2004:187). The second acquisitions register also displays alternating English and Afrikaans entries without any explanation or apparent policy. From 1964 when it became an independent cultural history museum, the registers and other internal records such as the catalogue cards were written in Afrikaans, which was one of the two official national languages of South Africa up to 1994.

With the new democratic government in South Africa, major policy changes were implemented. South Africa now has 11 official languages and museums are reviewing their language policies accordingly, which has an effect on cataloguing and indexing policies. If the cataloguing language is changed, measures to link old records to the new should be implemented. This may be done through translations of records or by developing multilingual thesauri to be applied for indexing and retrieval.

Levels of Specificity

The level of specificity should be appropriate to the collection, for example for a collection containing three clocks among other types of objects, the term 'clock' is a useful object name. However, for a specialist collection of clocks, more specific object names are required to distinguish between different types of clocks.

Very general object names can be assigned in order to group types of objects together, although if a structured thesaurus is used then this will be done automatically: for example the term 'furniture' would automatically include the term 'chair' as a narrower term. If a thesaurus is not used, both terms could be recorded separately (*CIDOC guidelines* 1995).

Vocabulary

It is important to use the same object name for all similar objects in a collection (*CIDOC guidelines* 1995). The names of objects and other subject indexing terms used for retrieval need to be standardised by using a controlled vocabulary list of some kind, preferably a thesaurus that gives a full structure of relationships between preferred and non-preferred terms, broader terms (BT) and narrower terms (NT). 'Beds' could then be retrieved in a search for all kinds of furniture (Will 1993:159).

No single thesaurus is generally accepted in the museum world. Various term lists have, however, been used or developed by different museums. Published lists include *Nomenclature* and the *Hertfordshire simple name list*, neither of which has a BT/NT relationship, and the *Art and architecture thesaurus*, which does not allow a term to have more than one broader term (Will 1993:159).

The *Art and architecture thesaurus (AAT)* developed by the Getty Information Institute is a carefully constructed terminology resource which focuses on more than just art and architecture. It provides terminology for the description, documentation and retrieval of cultural heritage information. Although its current focus is on the cultural heritage of Western Europe and North America, the Getty Information Institute Vocabulary Program is expanding its coverage. Terminology for African styles, periods and object names are being added (Lanzi 1998:30-31). The AAT is a very large list, containing 90 000 terms, of which 24 500 are descriptors. These are grouped into seven facets, which contain a total of 33 sub-facets or hierarchies (Will 1997:38). The seven facets are as follows: Associated concepts; Physical attributes (e.g. design elements, colours); Styles and periods; Agents (people, organisations); Activities; Materials; Objects.

Nomenclature's system (Chenhall 1988) is used by a few museums in South Africa. This system is concerned with a formal, generic object term and a general classification for that object based on its use. Information that is normally carried in secondary data fields such as material, style, period, quantity and maker is not used in creating object terms. 'Chippendale mirror' and 'Barbie doll' are not acceptable object terms. Object terms are generic identifiers that indicate an artefact's original function without reference to secondary data. MIRROR and DOLL are object terms. In a subject index, these generic terms enable the researcher to draw finer comparisons by establishing a generic population. (Chenhall 1988:II-2). *Nomenclature* provides a limited set of acceptable terms to identify objects and practise some simple conventions, such as inverting object terms so that nouns

precede adjectives to force like objects to appear together alphabetically. Plural forms of object terms are avoided (Chenhall 1988:I-1). Besides being brought together by virtue of sharing object terms, collection pieces are related in the classifications of the *Nomenclature* hierarchy (See Figure 2 in the addendum). For the purposes of indexing and collection management, the classifications are most important after object terms (Chenhall 1988:II-2). *Nomenclature* contains useful features such as 'use' references and related terms and some acquire regarding scope of terms. (See Figure 3 in the addendum for an example of an alphabetical list of terms.)

Another term list which is used in South African museums is *Trefwoorde vir kultuurhistoriese voorwerpe* (Slabbert 1974) which uses a similar binomial object term system. However it has no broad object categories such as the *Nomenclature* hierarchies.

The definitions or scope notes found in some term lists and thesauri (such as the *Art and architecture thesaurus*) are invaluable to the indexer, especially when the indexer is not an expert in the field of the objects which have to be indexed. When presented with object name inconsistencies, for instance during upgrading of collection cataloguing and indexing, the indexer will have to study literature (hopefully with good illustrations), compare all old records in the museum, consult with researchers and view the objects in storage. Recently such a case presented itself in the National Cultural History Museum, where the most recent inventory for the ceramics collection listed 18 ceramic vessels for storing food substances under the term VETPOT (vessels for storing butter or fat). When the original catalogue records were traced (by acquisition number), the following array of terms were found in the object name fields on the various object records:

CAPE STONEWARE; VETPOT; VETVAT; POT 18de-19de eeu; **VAAS**, 18de eeu; **VAAS,** 1890 - Keulse vaas (Cologne); **MARTAVAAN VAAS POT** (Martaban Pot. Martaban is a port in lower Burma); **MARTAVAANVAAS,** 18de eeu.

The indexer has to clear up two matters. These are firstly to ascertain whether the recent inventory identifies the objects correctly (giving them all the same name VETPOT) and secondly to evaluate the old naming system on the catalogue records. It is clear that the original object names on the old catalogue cards were not applied according to a rigid naming system: provenance, form and function were all applied as basis for object names. In cases like these, the indexer needs the inputs of the appropriate curator or researcher.

After viewing these objects the researcher reported that although they are all food storage jars, they were not necessarily used for fat. Water, oil, butter, pickles and even opium were also transported in these vessels. According to their physical appearance and history, they were of two types, namely the Cologne stoneware and Oriental stoneware named after the port Martaban. The curator suggested that all these vessels be indexed under a general object term linked to its use. JAR, FOOD-STORAGE was selected by the indexer, using the *Nomenclature* of Chenhall which is at the moment being introduced and tested as naming instrument at the National Cultural History Museum. The other terms, namely 'Martaban pot' and 'Cologne vase' or 'pot', will be used as 'use' references in the (English) catalogue. The English and Afrikaans terms and references will be mapped in a bilingual thesaurus.

Other Challenges concerning Object Names

The indexer has to find solutions to index objects which are not properly identified, objects with parts and accessories, object fragments, objects which are presented in pairs or sets and objects which cannot be easily linked to a single classification because they were made to serve many general purposes. *Nomenclature* (Chenhall 1988:II-4 to II-6) suggests different solutions, for example

SAXOPHONE CASE	or	SAXOPHONE [Case]
GLOVE PAIR	or	GLOVE [Pair]
LAMP BASE	or	LAMP [Part - base]

Each institution is urged to create its own rules to handle these situations.

As well as naming objects and retrieving them by subject, many other entities in relation to museum objects have to be indexed, such as people, organisations, places, events and materials.

People

Names of people who have some association with the object should be indexed, such as makers, users and donors. These names may be personal or they may represent groups such as organisations, institutions and cultural groups.

Detailed rules for the forms of names to be used for people and organisations are given in the *Anglo-American cataloguing rules* (2nd ed), widely used in libraries. There is general acceptance that these can be used as guidelines for museum applications too, although many existing users apply simpler rules of their own. The rules for the choice of names applicable to books are not so easy to apply to objects, because names are not generally printed on the items themselves; they have to be obtained from accompanying documents or supplied from the cataloguer's own knowledge or from reference sources (Will 1993:160).

In history museums, much attention is usually given to biographical information: personal, family and organisational histories are always recorded when objects are collected. These form biographical data files. Name indexing may be done by creating links between biographical and object records. According to Will (1993:160), biographical records can take advantage of a thesaurus structure to hold links between related people and organisations, as well as references from non-preferred to preferred forms.

Another source of names which might be consulted by the indexer is the *Union list of artist names* (ULAN). The first edition of the ULAN (1995) was the product of a seven-year effort to merge artist and architect names and biographical data from nine Getty Trust projects and databases in a single reference. Although the geographical coverage of the ULAN is global, its focus is on the post-medieval artists of Western Europe and North America. The ULAN does not advocate a single preferred form of an artist's name, but presents a range of choices that reflect art historical usage. In the ULAN, all name forms (including pseudonyms, nicknames, and orthographic and linguistic variants) are clustered together in a single entry. Each entry also includes biographical data and bibliographic

citations referring to the artist. The ULAN can be used as a source for authorities, or as a data value standard in the documentation (cataloguing, indexing and description) of cultural heritage information (Lanzi 1998:36-39).

Geographical Entities

When indexing places, it is convenient to create a full hierarchical structure, so that, for example, an item which has been indexed with the term Inverness will be retrieved in a search for any Scottish items, assuming that the computer system can accept an enquiry of the form 'Scotland and all its narrower terms'. Such systems are not widespread at present, so that museums often have to use the more cumbersome approach of including every level of the hierarchy in every record catalogued. The above item might then have to be given the terms 'Europe', 'United Kingdom', 'Scotland', 'Highland region' and 'Inverness' rather than just the one most specific term (Will 1993:160).

A vocabulary which may be used for an index to place names is the *Getty Thesaurus of geographic names* (TGN) which is a structured, hierarchical vocabulary containing names and other information about geographical places. The scope of the TGN is global; all continents and all nations are represented. The TGN is a vocabulary, not an authority. However, each place recorded in the TGN contains a 'preferred name' or entry-form name. It is necessary to flag one name to represent that place in the hierarchy and in alphabetical lists. Since the audience for the TGN is international, the preferred name is the name commonly used in the language of the place (the 'vernacular name') (Lanzi 1998:39-40).

The geographical name changes which take place over time usually present some challenges for the indexer. Each museum has to formulate a policy on the system to use, that is either index under the most recent geographical name and make references from the previous name to the preferred recent name. Otherwise both old and new names for places may be used in indexing, depending on the historical association that the museum object has to the place. References to relate these places should be made. In South Africa, there are numerous examples of geographical name changes over the years but the phenomenon has accelerated since the new democratic government came into power in 1994. Local name changes are for example'Pietersburg' in 'Northern Province' which is now officially 'Polokwane' in 'Limpopo' province. Examples of regional changes are the four provinces Gauteng, Mpumalanga, Limpopo and North West which previously formed part of one province, namelyTransvaal.

Dates

It is important to index the date or the range of dates associated with the history of objects or specimen. Patricia Harpring of the Getty Vocabulary Program in the United States touches on the issue of display information (on the object record) versus the structuring of the same information for indexing. These issues often arise when one tries to apply existing rules to the automated environment, particularly when the rules were devised to accommodate a non-computerised environment such as a card catalogue. Harpring (1999:33-34) states:

A successful catalog of art and cultural heritage information should maintain a healthy balance between flexible standards and consistent rules. On the one hand, it must allow the expression of uncertainty and ambiguity where the discipline requires it, and it must accommodate differences in style between departments and institutions. On the other hand, it must apply rules consistently where it counts – namely for information that is indexed for retrieval.

The expression of dates presents a common example of this dichotomy. It is necessary to allow date information to be displayed in a way that accurately expresses ambiguity, yet is intelligible to the end-user; at the same time, this information must be formatted consistently to allow retrieval. It is not always true that a work of art was created in a given year, such as "1915". Dates for art and cultural artifacts are often uncertain or approximate, as with dates of creation such as "probably 1511" or "ca. 430 BCE". Sometimes the date is known only to the specificity of a decade or a century, for example, "between 1880 and 1890" or "17th century". [...]

Which rules might best serve this fluid situation? All of the above examples illustrate "fuzzy" dates, which are not clearly delimited by a firm set of precise numbers. Various standards and guidelines deal with this issue in different ways. Many mainly address how to achieve a consistent display of date information, but do not adequately consider how to retrieve it.

A ... practical method is laid out in the Guide to the Description of Architectural Drawings and Categories for the Description of Works of Art (CDWA). According to these guidelines, dates of creation or dates for people can be recorded 1) as text or "display dates", and 2) the display date is then indexed using two integers delimiting the beginning of a date span and the end of a date span. The indexing dates are estimated based on criteria suitable for the material at hand, and are generally not visible to the end-user.

The principles discussed above with reference to dates apply to other museum information as well.

Styles and periods, events, activities and materials

To fully exploit the wealth of information contained in museum collections, indexes should be set up to retrieve objects by styles, period names, cultural traditions, materials and other subjects. Many sources are available for application of controlled vocabulary in indexing. (See paragraph on availability of vocabularies.)

Classification systems

A complementary technique to the alphabetical subject indexing which has been discussed is the assignment of classification notations or codes to object records. Classification notations have the advantage that they allow subject lists to be produced under complex subject headings. These may be specific to the context of the collection; they allow physical

arrangement of objects in store in subject order, like shelf marks in a library; and they allow selection of a predetermined subset of a collection. On the other hand, they have the disadvantages that an alphabetical index to the classification has to be provided, and ideally a searcher needs to have a copy of the classification headings that are relevant (Will 1993:159).

A classification system shares many of the advantages and disadvantages of a thesaurus, and, indeed, the hierarchical relationships of a good thesaurus lead to an implied classification in any case. One classification system quite widely used in the United Kingdom is the Social History and Industrial Classification (SHIC) which uses as its primary principle of division the historical context in which objects were used. This gives a different perspective from the thesaurus systems which focus on what the objects actually are (Will 1993:159). The main classes are

- Community life
- Domestic and family life
- Personal life
- Working life

Availability of vocabularies

Apart from the vocabulary tools which have been mentioned above, the indexer will be able to trace many other useful terminologies, covering many different subject areas by accessing the following websites:

http://www.chin.gc.za Canadian Heritage Information Network

http://www.mda.org.uk MDA (Formerly the Museum Documentation Association, UK) (Access the resource "wordHoard".)

Free text

As well as controlled indexing terms and classification schemes, a third necessary approach is the indexing of the individual words in object description. Searching on the actual words used in the descriptions is helpful for identifying half-remembered items, and will retrieve by specific terms peculiar to individual objects, trade names, jargon, et cetera (Will 1993:160). On the other hand inconsistencies in description mean that the searcher has to think of all terms which might have been used, including synonyms, grammatical variants, and generic and specific terms. Many existing inventory descriptions are inadequate, and editing them will take a lot of work, generally requiring examination of the objects. When searching free text there is no guarantee that items retrieved represent all items of that type in the museum, because other items may have been differently described (Will 1993:160).

Conclusion

Cataloguing and indexing of museum objects form an important part of collections management in museums. First it enables retrieval of information which is necessary

for accountability and security of the collections. Secondly, it facilitates the retrieval of subject information needed for research, publication, education and exhibits. Exchange of information between museums is being promoted in this way and development of common projects regarding research and exhibitions become a possibility.

BIBLIOGRAPHY

Aitchison, J & Gilchrist, A. 1987. *Thesaurus construction: a practical manual.* 2nd ed. London: Aslib.

Cataloguing cultural objects: a guide to describing cultural objects and their images (Draft - May 2004). Editors: M Baca ... et al. Portland, Oregon: Visual Resources Association. Available: http://vraweb.org/CCOweb (accessed on 20 May 2004).

Chenhall, RG. 1988. *The revised nomenclature for museum cataloguing: a revised and expanded version of Robert G Chenhall's System for classifying man-made objects.* JR Blackaby, P Greeno and The Nomenclature Committee. Nashville, TN: AASLH.

CHIN (Canadian Heritage Information Network) website. Available: http://www.chin.gc.za (accessed on 20 May 2004).

CIDOC guidelines for museum object information. 1995. International Committee for Documentation of the International Council of Museums (CIDOC). Available: http://www.cidoc.icom.org (accessed on 28 January 2000).

Grobler, E. 2004. Collections management practices (historical, anthropological and archaeological collections) at the Transvaal Museum 1913-1964. Draft D Phil dissertation, University of Pretoria.

Handbook of standards: documenting African collections. 1996. Paris: ICOM.

Harpring, P. 1999. Can flexibility and consistency coexist? Issues in indexing, mapping, and displaying museum information. *Spectra*, 26(1): 33-35.

Immelman, HFL. 2004. Personal communication. [Dr HFL Immelman was a SAMADOC Committee Member and Co-author of the Draft Museum Data Standard in 1987].

Lanzi, E.1998. *Introduction to vocabularies: enhancing access to cultural heritage information.* [S. l.]: Getty Information Institute.

Manuel de normes: documentation des collections Africaines / Handbook of standards: documenting African collections. 1996. Paris: ICOM.

Museum documentation standard 3rd draft: Main schedule. 1987. Cape Town: The Southern African Museums Association.

Slabbert, L. 1974. *Trefwoorde vir kultuurhistoriese voorwerpe.* Pretoria: Nasionale Kultuurhistoriese Museum.

Smither, R. 1990. "Is Britannia a personality?": some questions arising while indexing the Imperial War Museum's collections. *The Indexer*, 17(1):7-11.

Stanley, JL. 1989. Documenting African material culture, in *Africana resources and collections: three decades of development and achievement. A Festschrift in honor of Hans Panofsky.* Edited by JW Witherell. Metuchen, NJ: Scarecrow Press:118-150.

Spectrum: the UK museum documentation standard. 1997. 2nd ed revised and edited by J Cowton. Cambridge: Museum Documentation Association.

Stone, SM. 1992. Documenting collections, in *Manual of curatorship: a guide to museum practice.* 2nd ed. Editorial Board: JMA Thompson [et al.]. Oxford: Butterworth:213-228.

Wentz, P.1995. Museum information systems: the case for computerization, in *Collections management.* Edited by A Fahy. London: Routledge.

Will, L. 1993. The indexing of museum objects. *The Indexer*, 18(3):157-160.

Will, L. 1997. Data structures and indexing for museum collections management, in *Knowledge organization for IR: proceedings of the Sixth International Study Conference on Classification Research, held at University College London, 16-18 June 1997.* The Hague: International Federation for Information and Documentation (FID):36-41.

27 ADDENDUM

Figure 1: List of fields for collections of the Humanities and of Natural Sciences.
(*Handbook of standards: documenting African collections*. Paris: ICOM, 1996. p17-18.)

1	**OBJECT MANAGEMENT**	
[All fields of object management are relevant to both Humanities (H) and Natural Sciences (NS)]		
1.1	Country	
1.2	Institution where the object is located	
1.3	Owner institution name	
1.4	Accession number	
1.5	Acquisition or accession method	
1.6	Acquisition or accession date	
1.7	Acquisition source	
1.8	Permanent location of object	
2	**OBJECT DESCRIPTION**	
2.1	Image	H, NS
2.2	Object category by form or function	H
2.3	Object category by technique	H
2.4	Specimen form	NS
2.5	Body part	NS
2.6	Sex	NS
2.7	Age or phase	NS
2.8	Classified name (taxonomic)	NS
2.9	Common name	NS
2.10	Object name	H
2.11	Local name	H, NS
2.12	Name in another language	H, NS
2.13	Title	H, NS
2.14	Material	H
2.15	Technique	H
2.16	Dimensions	H, NS
2.17	Physical description	H, NS
2.18	Content	H
2.19	Inscription	H
2.20	Condition	H, NS
3	**HISTORY OF THE OBJECT**	
	PRODUCTION [Fields are only relevant to H]	
3.1	Producer	
3.2	Place of production	
3.3	Group of production	

3.7 Use
3.8 User
3.9 Place of use
3.10 Group of use
3.11 Date of use

COLLECTION [All fields of Collection are relevant to both Humanities (H) and Natural Sciences (NS)]
3.12 Collection or excavation locality
3.13 Site geographical coordinates
3.14 Coordinates of the object within the site
3.15 Site reference or name
3.16 Type of site
3.17 Age
3.18 Geological period of the feature from which the object was collected
3.19 Environmental details
3.20 Group from which the object was collected
3.21 Collector
3.22 Expedition
3.23 Collection or excavation date
3.24 Collection or excavation method
3.25 Collection field number
3.26 Historical comments

4 DOCUMENTATION
Each museum should organize its documentation according to its needs but the references for this documentation should be standardized to facilitate research.

Figure 2: Hierarchical list of preferred terms

(Chenhall, Robert G. 1988. *The revised nomenclature for museum cataloging: a revised and expanded version of Robert G Chenhall's system for classifying man-made objects.* James R Blackaby, Patricia Greeno and The Nomenclature Committee. Nashville, TN: AASLH)

CHAPTER IV - HIERARCHICAL LIST OF PREFERRED TERMS
CATEGORY 4: TOOLS & EQUIPMENT FOR MATERIALS

CATEGORY 4: TOOLS & EQUIPMENT FOR MATERIALS

*AGRICULTURAL T&E
Avarrancator ... use PRUNER, TREE
BAG, FRUIT PICKING
BALER, COTTON
BALER, HAY
Bar, Planting ... use DIBBLE
BARROW, APPLE
Basket, Berry ... use BASKET, GATHERING
BASKET, COTTON
Basket, Egg-gathering ... use BASKET, GATHERING
Basket, Fruit-picking ... use BASKET, GATHERING
BASKET, GATHERING ...rt BOWL, GATHERING
Basket, Potato ... use BASKET, GATHERING
BASKET, WINNOWING

BEARDER, BARLEY
BELT, BINDER
BILLHOOK ... rt HOOK, PRUNING
BINDER, CORN
BINDER, GRAIN
Binder, Row ... use BINDER, CORN
Binder, Twine ... use BINDER, GRAIN
Binder, Wire ... use BINDER, GRAIN
BLOWER, ENSILAGE
Blower, Forage ... use BLOWER, ENSILAGE
Blower, Impeller ... use BLOWER, ENSILAGE
Blower, Silage ... use BLOWER, Ensilage
Blower, Silo ... use BLOWER, ENSILAGE
Board, Dibble ... use DIBBLE
Board, Spotting ... use DIBBLE
Boat, Stone ... use STONEBOAT

Figure 3: Alphabetical list of preferred terms

(Chenhall, Robert G. 1988. *The revised nomenclature for museum cataloging: a revised and expanded version of Robert G Chenhall's system for classifying man-made objects.* James R Blackaby, Patricia Greeno and The Nomenclature Committee. Nashville, TN: AASLH)

CHAPTER V - ALPHABETIC LIST OF PREFERRED TERMS

BAG, NEEDLEWORK	TEXTILEWORKING T&E	104
BAG, ORE	MINING & MINERAL HARVESTING	97
BAG, PAJAMA	PERSONAL GEAR	35
BAG, PASTRY	FOOD PROCESSING T&E	58
BAG, POWDER	ARMAMENT - ACCESSORY	130
BAG, PUNCHING	SPORTS EQUIPMENT	220
Bag, Sand	REGULATIVE & PROTECTIVE T&E	
use SANDBAG		
BAG, SCHOOL	PERSONAL GEAR	35
BAG, SLEEPING	BEDDING	9
BAG, TAMPING	MINING & MINERAL HARVESTING T&E	97
Bag, Tobacco	PERSONAL GEAR	
use POUCH, TOBACCO		
BAG, TRICK OR TREAT	CEREMONIAL ARTIFACT	206
BAG, WATER	LTE - ACCESSORY	193
BAG, WHALER'S	FISHING & TRAPPING T&E	56
BAGATELLE	GAME	217
BAGPIPE	MUSICAL T&E	172
BAILER	MINING & MINERAL HARVESTING T&E	97
BAILER	WATER TRANSPORTATION - ACCESSORY	201
BALALAIKA	MUSICAL T&E	172
BALANCE	WEIGHTS & MEASURES T&E	166
BALANCE, ANALYTICAL	WEIGHTS & MEASURES T&E	166
BALANCE, ASSAY	MINING & MINERAL HARVESTING T&E	97
BALANCE, BERANGER	WEIGHTS & MEASURES T&E	166
BALANCE, BOB	MINING & MINERAL HARVESTING T&E	97
BALANCE, DEMONSTRATION	WEIGHTS & MEASURES T&E	166
BALANCE, GRAM-CHAIN	WEIGHTS & MEASURES T&E	166

BALANCE, ROBERVAL	WEIGHTS & MEASURES T&E	166
BALANCE, SPECIFIC-GRAVITY	WEIGHTS & MEASURES T&E	166
BALANCE, TORSION	WEIGHTS & MEASURES T&E	166
BALANCE, WATER	MINING & MINERAL HARVESTING T&E	97
BALCONY	BUILDING	1
BALDACHIN	CEREMONIAL ARTIFACT	206
BALDRIC	ARMAMENT --- ACCESSORY	130
BALE	CONTAINER	186
BALER, COTTON	AGRICULTURAL T&E	41
BALER, HAY	AGRICULTURAL T&E	41
BALL	TOY	223
BALL, BILLIARD	SPORTS EQUIPMENT	220
BALL, BOWLING	SPORTS EQUIPMENT	220
BALL, CROQUET	SPORTS EQUIPMENT	220

28 NATIONAL BIBLIOGRAPHIES
A Lesotho Case Study

David Ambrose

Abstract

Early Lesotho bibliography was encompassed within volumes which covered South Africa, but generally omitted items in African languages. An account is given of the origins of a book published in 1980 which attempted to survey with annotations all published literature on Lesotho. The subsequent Lesotho Annotated Bibliography evolved from conventionally published format to a computer database, whose structure and arrangement is described. Details are given of the persons most directly involved in these activities, and the current challenges to make the bibliographic database more comprehensive and accessible. Indexing is discussed in relation to electronic media, hard copy format, and making provision for merging a number of separate indexes. Problems arising from Sesotho orthography and place names are discussed.

Introduction

Many African states have boundaries which were defined by colonial powers. Lesotho, however, has a rather different history. It was already a nineteenth century kingdom, when its monarch King Moshoeshoe sought British protection to avoid conquest from the neighbouring Orange Free State Republic. For 98 years from 1868, Lesotho (known to the British as Basutoland) was under British rule, until independence was restored to the Crown Colony in 1966. (Contrary to the belief of many, its legal status was not that of a protectorate.) The country is unusual in Africa in having a single national language, Sesotho, which shares with English the status of being one of the official languages. It also has the distinction of being an enclave within a larger state, the Republic of South Africa. This is a consequence of the people of Lesotho, the Basotho, specifically requesting Britain to exclude them when four former British colonies became the Union of South Africa in 1910. The Basotho had had a disastrous period (1871-1883) when the British had annexed Basutoland to the Cape Colony. In a period when the franchise excluded black voters, they did not again want to be ruled by a Parliament in which they would not have a voice. There is one other curious feature about Lesotho. Almost all of the early literature about the country is not in either of the official languages but in French, the language of the two main groups of missionaries, the Paris Evangelical Missionary Society (PEMS) and the Oblates of Mary Immaculate, who first established missions, schools and printing works in Lesotho. The PEMS arrived in Lesotho in 1833 and the Catholic Oblates in 1862.

Early Bibliography

As a result of its history and geography, Lesotho occupies an equivocal position bibliographically vis-à-vis South Africa. Early bibliographers such as Mendelssohn (1910) tended to regard Lesotho as an integral part of South Africa. The revision and continuation of Mendelssohn's work *A South African bibliography to the year 1925* (1979), together with its supplement (Rossouw & Vockerodt 1991) and indexes (Rossouw 1997), provided good coverage of many items published in or about Lesotho up to 1925, but specifically excluded government publications, newspapers, maps and African language items. As a result, the then rapidly growing Sesotho literature was excluded.

Sesotho literature and newspapers to 1907 were fairly comprehensively covered in a manuscript catalogue finally published 80 years later (Ellenberger 1987) and a useful listing of Sesotho books up to 1965 was made by Ferragne (1974). For newspapers, an annotated bibliographic guide covering the period to 1976 was made by Switzer and Switzer (1979). There are a number of other bibliographies exclusively or partially devoted to Lesotho or Sesotho which are listed in Willet and Ambrose (1980). Several of these were bibliographies undertaken by students at the University of Cape Town School of Librarianship and the University of the Witwatersrand Department of Bibliography, Librarianship and Typography over the period 1946 to 1970.

The 'Comprehensive' Bibliography

Tertiary education in Lesotho had begun as early as 1945 with the establishment of Pius XII College at Roma. Its assets were acquired by the British Government in 1964 to establish the Roma Campus of what was then known as the University of Basutoland, Bechuanaland Protectorate and Swaziland (UBBS) (renamed in October 1966, when two of the countries it served gained independence, as the University of Botswana, Lesotho and Swaziland [UBLS]).

Among staff recruited to UBBS in 1965 were Shelagh Willet as Deputy Librarian and David Ambrose as Lecturer in Mathematics. Ambrose developed an interest in Lesotho literature to the extent of making a personal collection and keeping an annotated card index of other items about the country which he located in libraries and collections in the region and overseas. Willet in the meantime founded in the University Library the Boleswa (ie Botswana, Lesotho & Swaziland). Collection of materials specially devoted to the three countries which UBLS served.

When an offer came to Willet from the Clio Press to compile an annotated bibliography for Lesotho in their planned World Bibliographic Series, she approached Ambrose, resulting in a collaboration of more than 10 years which led finally to the publication of the volume *Lesotho: a comprehensive bibliography* (Willet & Ambrose 1980). The volume differed from others in the series, which were select bibliographies, because Willet and Ambrose persuaded Clio as an experiment to allow them to compile a 'comprehensive' bibliography, arguing that the total literature in Lesotho as a relatively small country could be encompassed within a single volume, even if the book included theses, periodical articles, newspapers and what was then a relatively manageable total corpus of 'grey literature'. (At

the time desktop publishing was virtually unknown, photocopiers were still relatively rare, and the consultancy business was still in its infancy, so that there was comparatively little informally published grey literature.)

The method of compilation at the time was on cards, which included both annotations and keywords; keywords were divided into primary and second keywords according to importance, although in the final product these keywords were detached from their entries and conflated without distinction into a massive 76-page computer-generated index which also included authors and book titles (although not titles of articles in periodicals).

When they took it on neither Clio nor the compilers realised how much would be involved in a comprehensive bibliography. The final product, which included an extensive publishing history for Lesotho was a book of over 500 pages and 2 562 individually annotated items (Willet & Ambrose 1980). The book was for a long time the largest in the Clio World Bibliographic Series (larger than those which followed on the United States and on the Soviet Union!) and ultimately, among bibliographies for individual nations, only the World Bibliographic Series volume on Japan was larger.

Eventually Clio required a revised volume, and for this conformity to the select bibliography nature of the rest of the series was a requirement. The bibliography, as will be seen, had meanwhile grown as a comprehensive bibliographic database, so it was left to someone else to produce the select bibliography (Johnston 1996), which was reduced to 563 entries (a number of them taken from the original work unchanged) and well over 200 pages. On this occasion, separate author (5 pp.), title (15 pp.) and subject (9 pp.) indexes were provided. Another change was that within sections, entries were alphabetical by title rather than by author as in the original work.

Bibliographic Activities in the 1980s

Because of its size, the bibliography published in 1980 took three years to emerge, and hence only included items up to 1977. Meanwhile the circumstances of its compilers had changed.

Willet had become the founder librarian of the Botswana Campus of UBLS in 1971, but for a number of years left librarianship to manage a Quaker-funded centre for refugees in Botswana. She ultimately returned to bibliography and participated in an ongoing series of volumes on the Khoe and San (Willet, Monageng, Saugestad & Hermans 2002; Willet 2003).

Ambrose meanwhile was seconded for the period 1981 to 1985 from the Department of Mathematics at what was now the Roma Campus of the National University of Lesotho to become founder of the Documentation and Publications Division of the university's recently created Institute of Southern African Studies. This gave him the opportunity to set up a publicly available collection of many of the items in the bibliography, together with new items which included many examples of the now rapidly expanding grey literature genre. He also initiated the publishing of bibliographic supplements to the comprehensive bibliography in seven issues of a periodical called *Lesothana* (1982-1984). These published issues both covered new materials, but also added many older items to the original 'comprehensive' bibliography, which had certainly been less than comprehensive in relation to articles in nineteenth century French missionary periodicals. The supplements

covered 700 entries and some 35 per cent of the areas in the 1980 work when Ambrose was reclaimed by the Department of Mathematics which suffered perennial staff shortages.

Bibliographic Activities in the 1990s and Beyond

In 1990, Ambrose was able to negotiate a new status as Honorary Associate Research Professor attached to the National University of Lesotho's Institute of Education. This unpaid position gave him considerable flexibility to undertake various activities, a number of which resulted in publications in a variety of fields. A few of these activities required bibliographic compilations of which the most notable was an annotated 320-page bibliography for Lesotho water resources management (Ambrose 1996). At the same time, books that had appeared since the 1980 bibliography were systematically listed alphabetically by author in successive issues of the periodical *NUL Journal of Research* (1991-). By 2001 (the most recent issue), there had been nine issues of the periodical and a total of 475 Lesotho-relevant books had been listed, covering authors with names from A to the first part of the letter M.

By the mid-1990s, the Lesotho bibliography database was fragmented between the 1980 published volume, the various published items just mentioned and a card index of over 3 000 cards which had been accumulated since the 1977 cut-off date for the 1980 published book. To facilitate retrieval of items by topic, these cards had all been photocopied and made into a second card index using the same topic headings of the 1980 book. The bibliographic work was however at an important turning point. The amount of material was too great to be published except in a series of volumes, and such a series would in any case not be commercially viable and would soon be out of date.

Up to this point, computers had been used as word processors to facilitate publication, but not for a main database. A computer database for the *The Lesotho annotated bibliography* was first created in 1996, and subsequently the aim has been to merge all earlier activities as well as new items into a single accessible database. The details can be found in the fourth edition of *The Lesotho annotated bibliography: a brief survey* (Ambrose:2003a). As at July 2003, the database was divided into 24 'volumes' subdivided overall into a total of 196 sections, each covering a subject area and each section prefaced (although not all prefaces had yet been written) by a summary of the subject area and its known Lesotho literature. Since many items have content relevant to several sections, it was decided at an early stage to include cross-references between sections. The sections themselves are arranged in orders appropriate to their content. For example, whereas the section 'Fiction in Sesotho' is logically arranged alphabetically by author, the section 'Census Reports' is arranged chronologically. A few sections are also arranged numerically, for example the section '1: 50 000 & 1: 100 000 Scale Maps' is arranged by map sheet number, while for the section 'Lesotho Highland Water Project Contracts' the contract numbers form a logical ordering system. Overall, out of 196 sections, 117 are arranged chronologically, 71 alphabetically and 8 numerically.

As at July 2003, the 196 sections contained a total of 4 882 items, 1 848 cross-references between sections and 3 168 pages of text. Priority was given to putting new items into the database. Consequently transfers from the various complementary bibliographic sources

were still some way from completion, with only 13 per cent of *Lesothana* items, 61 per cent of *NUL Journal of Research* items, 64 per cent of the water resources bibliography (Ambrose 1996), and 18 per cent of *Lesotho: a comprehensive bibliography* (Willet & Ambrose 1980) transferred. These were deemed less urgent because of their availability in another published form. However, it was of some concern that at this point only 41 per cent of the 3 492 unpublished cards covering the period 1977-1996 had so far made it into the bibliography.

The computer database as a whole is not online or widely available. Its appropriate publication may be on CD-ROM, when it is more complete. There are two public access points in Lesotho via Sechaba Consultants in Maseru and the Campus Bureau in Roma. However, sections of the bibliography for which there is a high demand are available published in hard copy, each with an appropriate ISBN; 14 such sections are available so far, of which three have reached second editions.

Indexing

The aim of indexing is to provide the user of a book, periodical or other media item fast access to information. Technological developments have both facilitated indexing and at the same time posed questions as to the need for traditional indexes of the type commonly found at the back of a publication. If the item is available in electronic format, a search command with more flexible alternatives than any traditional index can find information very quickly.

In relation to the bibliographic database, the practice of including keywords at two levels has been maintained, so that each item may have principal keywords indicating that for these topics it is an important source of information, and secondary keywords indicating it as a significant but less important source of information on those topics. The principal keywords are distinguished by using boldface type. However, although this has been the continuing convention with the Lesotho bibliographic work, in practice the indexes themselves have not so far distinguished between the two levels of keywords.

The present practice has been to provide indexes only to the published sections of the bibliography. These are compiled in an initial format where each entry is prefixed by a dot, and the 'dot' index stored for future use. For example, taking Section 132, *Rock paintings* (Ambrose 2003b), there are two alphabetical indexes, one for authors, compilers, editors and artists and the other for keywords. An author entry in the dot index may read 'Inskeep, R R .320' in the initial dot index which is stored, but reads 'Inskeep, R R 320' in the actual printed index. The full bibliography number for the item is 132.320, but within the section, since all items have the prefix 132, it is unnecessary to repeat them. Thus all dots are deleted by a simple exchange command to create the index for the published section. However, when it comes to merging indexes a different exchange command, replacing '.' by '132.' is used, so that the full item number appears and can be merged into an index with items from other sections of the bibliography. At this point, merged indexes have not been generated, but it has seemed important to provide for this possibility.

Any Lesotho index is likely to contain a few Sesotho words or names which can create sorting problems or problems of consistency. Diacritics are illustrated by the name *'Mantšebo*, which conforms to the standard 1906 Sesotho orthography used in Lesotho,

although in South African Sesotho orthography would be written *Mmantshebo*. The *tš* provides no problem, because computer sorting usually assigns it to a position immediately after *ts*. However the double M – written *'M* – will consign such words to the very front of any alphabetical listing whereas they are normally to be sought alphabetically under M. In practice, removal of the initial diacritic before sorting, and restoring it afterwards has been used to achieve the expected alphabetical sequence. The same remarks apply to the double N – written *'N* – which is of rather rarer occurrence.

Place names constitute another problem. A place may be called *Ha Ramabanta* which is equivalent to 'Ramabanta's place', usually written 'Ramabanta's' in English, but in many cases shortened to 'Ramabanta'. In such cases, the indexing word is taken to be *Ramabanta, Ha* or *Ramabanta*, if this seems to be the commonest form in use. The one exception to this is the best-known rock paintings site in Lesotho, *Ha Baroana*, which is not a village name, but means 'place of the little Bushmen'. This name is never found in any other form, and is therefore indexed under H not B.

The Future

The Lesotho annotated bibliography has a rather precarious future, because of its creation through individual efforts. Even if it had been an officially sponsored activity it is by no means certain that it might have endured as long, given the chequered, indeed rather sad history of such institutions as the National Library, National Archives and National Museum in Lesotho. The bibliography's usefulness is not in doubt, because sales of hard copies to individuals and libraries are steady, and the electronic database is searched, despite a fee, by many persons undertaking research. In many cases it facilitates a literature search, at least in relation to Lesotho items, for almost any research topic, although it very often provides only a single known location for any particular item, the place where it was in fact seen, and this is often the compiler's own 'House 9' collection. Obviously additional locations ought to be sought, a time-consuming exercise, and the very many items not yet entered need to be added. The problem is that the rate at which new literature, particularly in the 'grey literature' category, is emerging makes it increasingly difficult to keep the database reasonably comprehensive. It currently has strengths in subject areas where most documents are included, and weaknesses in subject areas where items, although often collected and sitting in 'pending' boxes, may at this point be as little as 10 per cent entered.

No simple solution in relation to the future of the bibliography is offered at present, but it is noted that for the present writer, the bibliography has been a hobby (albeit one which has become almost an obsession) and has had to take its place alongside many other activities. He has now reached retirement age, and finding persons interested in bibliographic work able to write appropriate annotations is not easy, the more so when the rewards are not monetary but rather in the realm of intellectual satisfaction.

BIBLIOGRAPHY

Ambrose, D. 1996. Annotated bibliography for Lesotho water resources management, in *Water resources management: policy and strategies.* Maseru: TAMS Consultants for Lesotho Government Ministry of Natural Resources Department of Water Affairs.

Ambrose, D. 2003a. *The Lesotho annotated bibliography: a brief survey*. 4th ed. Roma, Lesotho: House 9 Publications.

Ambrose, D. 2003b. *Rock paintings.* 2nd ed. Roma, Lesotho: House 9 Publications. (Lesotho Annotated Bibliography Section 132).

Ellenberger, DF. 1987. *Catalogue of the Masitise Archives.* Roma, Lesotho: Institute of Southern African Studies & Morija, Lesotho: Morija Archives.

Ferragne, M. 1974. *A catalogue of 1000 Sesotho books.* Roma, Lesotho: Social Centre.

Johnson, D. 1996. *Lesotho.* Rev ed. Oxford: Clio.

Lesothana: an annotated bibliography of new and newly located Lesotho materials. 1982 - 1984. Roma, Lesotho: National University of Lesotho Institute of Southern African Studies. Documentation Centre (7 issues, quarterly, August 1982 - February 1984).

Mendelssohn, S. 1910. *Mendelssohn's South African bibliography* . London: Kegan Paul,Trench, Trübner.

NUL Journal of Research. (1991-). Roma, Lesotho: National University of Lesotho.

Rossouw, F (comp). 1997. *A South African bibliography to the year* 1925, vol. 6: Subject and title indexes. Cape Town: South African Library.

Rossouw, F & Vockerodt, EA (comp). 1991. *A South African bibliography to the year 1925, vol. 5: supplement.* Cape Town: South African Library.

A South African bibliography to the year 1925. 1979. London: Mansell. 4 vols.

Switzer, L & Switzer, D. 1979. *The Black Press in South Africa and Lesotho: a descriptive bibliographic guide to African, Coloured and Indian newspapers, newsletters and magazines 1836-1976.* Boston, Mass: Hall.

Willet, S. 2003. *The Khoe and San: an annotated bibliography, volume 2.* Gaborone: Lightbooks.

Willet, S, Monageng, S, Saugestad, S, & Hermans, J. 2002. *The Khoe and San: an annotated bibliography, volume 1.* Gaborone: Lightbooks.

Willet, SM & Ambrose, DP. 1980. *Lesotho: a comprehensive bibliography.* Oxford: Clio.

29 NEWSPAPERS

Christopher Merrett

Abstract

This chapter considers the special features associated with the indexing of newspapers and provides guidelines for the indexer. The example of indexing *The Times* of London is used as many common features are applicable. The chapter concludes by examining developments in online indexing for easier information retrieval.

Introduction

Although the accuracy of the information they contain must always be questioned and treated with caution, newspapers are an indispensable source of primary historical evidence. Among other important roles their information gathering provides support for civil society and their very existence is an emblem of the crucial importance of the freedoms of information and expression to a democratic society. It is no accident that they are often referred to as the Fourth Estate. Titles that have a reputation for the quality of their reportage are assigned long-term status as 'newspapers of record'. Others might have a distinctive role to play at specific historical periods because of the nature of their sources. For instance, certain progressive South African newspapers had particularly close links with the Mass Democratic Movement during the terminal years of apartheid. Newspapers can thus be viewed as both an academic tool and a civil society resource. Their indexing might accordingly span particular historical periods or be current, depending upon circumstances.

One of the world's most comprehensive newspaper indexes is that of the London newspaper, *The Times*, whose scope and presentation provide a benchmark for all other endeavours in the field. The 2002 edition runs to over 1 500 closely printed pages and its content reflects the *Sunday Times* as well as the well-known supplements such as the *Times Literary Supplement*. Because of its pre-eminent status, reference will be made to the indexing practice of *The Times* in this contribution.

Indexing Features

Perhaps the most challenging feature of newspapers is that 'there is no standard way to index' them (Beare 1999:24). They are documents that reflect the societies in which they are published and their indexes must 'fit the time, the place and the occasion' (Friedman 1942:23). A true index to newspaper articles is a complex matter of detail and precision

(Semonche 1986:27): *The Times* arranges entries chronologically, giving the page number and column (a-h), plus an indication of short (s) articles with a column length less than 50 centimetres and the presence of photographs, maps and charts. Most non-commercial projects should more accurately be described as selective subject indexes.

The indexing of newspapers, while a relatively common activity, has a number of features that pose a distinctive challenge to the indexer. Before starting a project it is important to establish broad parameters. These will depend on two main considerations: the probable needs of the users for whom the index is being compiled, and the time and resources available to the indexer. For these reasons very few newspapers will be indexed in their entirety, making this an exercise completely distinct from, for example, the indexing of a book where every part of the text (and sometimes even footnotes and endnotes) is taken into consideration. Selectivity and the initial decision about what to omit are crucial (Sandlin, Schlessinger & Schlessinger 1985:186). For instance, a typical policy would be to restrict indexing to the news content only, ignoring everything else. On the other hand, a retrospective project might conceivably involve the indexing of advertisements and the births, marriages and deaths columns for the value of their historical evidence. Such decisions need to be made at the outset: backtracking can be a costly exercise. For this reason a pilot index is probably a sensible option, if feasible.

Another feature of current newspaper indexing that sets it apart from other genres is the fact that the indexer is dealing with an evolving story, sometimes spanning a lengthy period of time. At the outset it is often not possible to isolate the key concepts that will provide the crucial descriptors (Coates 1983:184; Kyte 1967:127). Sometimes a retrospective view will clarify such matters, but in other instances even time will be of no assistance and the story will fade into obscurity. What exactly, to use a South African example, was the McBride affair of 1998 about? Was his arrest and detention in Mozambique a story about a maverick government official; relations between two sovereign states; or dirty tricks by remnants of the old regime? A more recent example, concerning events around the former Director of Public Prosecutions, Bulelani Ngcuka, raises comparable questions about infighting among ANC members, corruption around the arms deal, or the settling of apartheid-era scores. In both cases the indexer may have little option but to use descriptors based on individual and corporate names. South African stories that originally appear to be about corruption often turn out to have other dimensions.

Guidelines

The golden rule of any indexing is to rely on a thorough reading of the text, but with newspapers there is a strong temptation to be influenced by headlines. These are designed to be eye-catching for the reader rather than an accurate reflection of content and are not to be relied upon on their own, although they may be useful as indicators. As Ahmed (1991:257) emphasises, the indexer depends on content plus background knowledge to identify the main significance of a story. The good newspaper indexer needs to be well informed, a quality that is of course reinforced and enhanced by newspaper indexing in a self-reinforcing process.

Newspaper indexers find that in order to avoid over-indexing (Beare 1989:229) it is sensible to apply a rule of thumb regarding the number of descriptors that might be applied to any one article. Some will be adequately covered by one; others may invite many. It is useful practice to regard the first descriptor assigned as covering the main theme of the article, with up to three or four others supplementing it. It is doubtful if a standard length article would require more than five descriptors (*The Times* uses a maximum of four, as specific as possible). Apart from keeping their numbers within realistic limits, it is essential to maintain consistency from issue to issue (Perica 1975:3), remembering that the indexer is involved in an evolving exercise, not a finite text. Although it is possible to use a standard, published general list of descriptors as a basis for indexing, the indexer will in all likelihood develop a thesaurus of her or his own in due course. A long-term project indexing the *Mail & Guardian* (*M&G Index*) of Johannesburg unexpectedly led to compilation of just such a thesaurus that has been used to describe other types of published material about South Africa held in libraries (Merrett 2000:73). But with the passage of time it will inevitably be necessary to engage in a certain amount of retrospection and, if necessary, adjustment (Knee 1982:101). Newspapers tend to have a primary geographical context within which they work and this can become the default for the descriptors: *The Times* presumes that the country is Britain, unless otherwise stated.

Given that a newspaper has the potential to throw up many indexable personal and collective or corporate names, clear guidelines have to be established. Names should generally only be used when they are clearly the primary focus of an article (although letters to *The Times* are indexed under both the writer and their subjects). In general the index user is far better served by an emphasis on concepts. An interesting exception was in fact provided by the *M&G Index* already mentioned: when it was started, under conditions of considerable repression, it was important to use the names of organisations in order to provide access to their otherwise censored opinions and policies. But as democracy asserted itself, so it was possible to change to an indexing practice that emphasises concepts. This helps to maintain long-term indexing consistency (Einhorn 1976:2-3).

Another important consideration lies in the construction of descriptors. Natural language is preferable for the sake of the user: Traditional healers is more straightforward a term than Healers, traditional – which will inevitably require a *see* reference. But there will be cases where the grouping of headings is useful (Christie 1986:92). In the *M&G Index*, whose main focus is the political, social and economic condition of South Africa, the following collective headings are used: Broadcasting; Children; Churches; Courts; Farms; Schools; Taxes; Violence; Women; and Workers. 'Obituaries' is also a collective descriptor, but individual names are indexed as well. *The Times* is extremely sparing with collective entries and limits them, surprisingly perhaps, to leading articles, law reports, obituaries and editorial cartoons. Grouping is not to be confused with subdivision, which is possible with any heading for the sake of clarity and economy of space. For instance:

```
Freedom
        of association
        of expression
        of the press
```

Cross-references (*see* and *see also* references) should be used with intelligent and sparing anticipation and chosen for their potential usefulness. The most readily understood alphabetical order is word by word (practice at *The Times*).

Using natural language consistently with a majority of direct descriptors and a modicum that is collective or grouped will create a long-term structure or overview of a newspaper that gives it added identity (Dewe 1972:58). All indexers create a resource that is greater than the sum of its parts. By its very nature a well-structured index will stimulate creative research, using newspapers by highlighting prominent topics or even encouraging serendipity.

Online Developments

Most major newspapers can now be consulted online, some by using sophisticated search programs. The ability to scour enormous volumes of text by keyword suggests that old-fashioned indexing is now redundant. Under certain circumstances this might be true. But while recall potential will inevitably be high, precision is another matter. To take an obvious example, the search term 'Mandela' will score a multitude of hits in any South African newspaper database: by searching the *M&G index* online for the 1999 issues, 222 hits were derived, but in the opinion of the searcher only three of these were of major significance (Merrett 2000:74). The search itself will not reveal which are the particularly significant hits. Human agency in the form of the indexer is alone responsible for this. Such a view is forcefully put by Steemson (1994:21) writing about the *Express Newspapers* archive in London:

> We chose human brainpower for entering index references rather than the cheaper option of full-text automation. No full-text system invented can emulate the trained selectivity of the human mind, or the creative intuition of an experienced librarian.

Conclusion

Of course many people still do not have access to online search facilities, nor are these necessarily quicker to use than a well-constructed manual index. The manual index will undoubtedly be more adept at picking up and highlighting significant articles and key information: indexing cannot be reduced to a mechanical exercise. One of the reasons for this is that descriptors often need to be contextualised.

BIBLIOGRAPHY

Ahmed, N. 1991. Newspaper indexing: an international overview. *The Indexer*, 17(4):257-267.

Beare, G. 1989. Local newspaper indexing projects and products. *The Indexer*, 16(4):227-233.

Beare, G. 1999. *Indexing newspapers, magazines and other periodicals*. Sheffield: Society of Indexers.

Christie, P. 1982. Indexing 19th century North Devon newspapers. *The Indexer*, 15(3):91-93.

Coates, PR. 1983. The retrospective indexing of two colonial newspapers. *The Indexer*, 13(3):183-184.

Dewe, M. 1972. Indexing local newspapers. *Assistant Librarian*, 65(4):58-59.
Einhorn, JM. 1976. *Guidelines for indexing local newspapers.* Westerly, RI: Westerly Public Library.
Friedman, HA. 1942. *Newspaper indexing.* Milwaukee: Marquette University Press.
Knee, M. 1982. Producing a local newspaper index. *The Indexer*, 13(2):101-103.
Kyte, C. 1967. The Times index. *The Indexer*, 5(3):125-127.
Merrett, C. 2000. Yesterday is history: the evolution of the Weekly Mail index. *The Indexer*, 22(2):72-75.
Perica, E. 1975. *Newspaper indexing for historical societies, colleges and schools.* New York: Library Research Associates.
Sandlin, LM, Schlessinger, J*H & Schlessinger, BS. 1985. Indexing of smaller-circulation daily newspapers.* The Indexer, 14(3):184-188.
Semonche, BP. 1986. Newspaper indexing: in search of a solution. *Collection Building*, 7(4):24-28.
Steemson, M. 1994. Direct electronic access to a large clippings library. The Indexer, 19(1):19-21.
The Times index. Reading: Gale.

30 POLITICS AND INTERNATIONAL RELATIONS

Jacqueline A Kalley and Elna Schoeman

Abstract

The chapter commences with definitions of the closely related disciplines of politics and international relations, prior to focusing on their unique features and indexing characteristics. Guidelines demonstrate concrete methods of approaching the indexing tasks. These are illustrated by many practical examples. Two case studies of libraries relevant to the topic provide an insight into the practical problems encountered.

Introduction

The dynamics of both politics and international relations are twofold. On the one hand, they are driven by the regularity of G8 summit meetings or elections at five-yearly intervals, to quote some obvious examples. On the other hand they are subject to the contingencies of the spontaneous, the unknown or the unexpected. Although the United States was fully aware of the possibility of terrorist attacks, who could have predicted the 9/11 attack on the Twin Towers and the Pentagon? The implications have changed the face of both politics and international relations forever. It is the role of the indexer to document these occurrences in a way that facilitates the quick and easy retrieval of pertinent information.

The study of politics and the relatively new discipline of international relations falls firmly into the social science discipline. The former has been defined as 'the science or study of government, and the interaction of political forces, with reference to its principles, aims, methods and conduct' (Comfort 1995:464). This rather bald definition could be amplified by an explanation devised by Benyon and Jeffery (2000:1) in a lecture celebrating the first 50 years of the Political Science Association in Great Britain and excerpted below:

> So what is it that links the diverse subject matter of political studies, and why should we explore it? One definition is that it is "an activity which involves conflict and disagreement over ideas and interests and its resolution through co-operation, compromise or coercion, and as such it entails institutions, rules and procedures". More brutally, perhaps, it has been called "a strife of interests masquerading as a contest of principles". There is also an important role for political studies in the current debate about the relationship between citizens and the state. At the same time, the demands for, and impact of, research by politics academics has grown greatly... New ideas and perspectives have been opened up in fields as crucial as the post-Cold War era, local governance, and constitutional reform. Governments

have been engaged both in designing the apparatus for new institutions and electoral systems, and in evaluating their operation.

These sentiments could well be applied to the Southern African region since they indicate the emergence of new concepts as the political landscape changes and evolves. For the purposes of this chapter, only one area of politics has been selected in order to clarify indexing techniques.

Politics

Elections have been chosen as they constitute a specialised branch of politics; and indeed there are libraries both in South Africa and abroad devoted solely to the topic. Elections have been described as the 'actual process of choice, comprising the campaign, the count, and the declaration'. To Stanley Baldwin (Comfort 1995:177) they were a regrettable necessity: 'I hate elections but you have to have them. They are medicine.' Love them or hate them, transparent and credible elections have become the benchmark of democracy and their execution and evaluation have become a highly complex activity in determining whether the government stays in office in the main house of the legislature or is replaced by a new administration led by a different party. The models of electoral systems, electoral administration, legislation, the logistics of the polling day and observation all constitute vital elements in the successful operation of an election and this is reflected in the indexing of the material both surrounding and emanating from them.

International Relations

The Penguin dictionary of international relations defines it as follows: 'The term … is used to identify all interactions between state-based actors across state boundaries' (Evans & Newnham 1998:274). International economic relations, international politics and international law are related fields but the methodology differs because that of international relations is eclectic, utilising methods of study from various fields. The discipline of international relations is a multidisciplinary and heterogeneous field of study related to those mentioned above. Although previously studied within the parameters of history and politics, it was given formal recognition in 1919 when the first chair was established at the University of Wales, Aberystwyth.

As such, these fields of study should be relatively easy to index because one would apply the well-defined indexing rules; however their unique characteristics make these subjects rather more challenging than one would expect.

Unique Features

- Both politics and international relations are policy oriented.
- Theoretical directions include the concepts of idealism, and have expressed a belief in progress towards a fundamentally peaceful and just world order.

- Subsequent theoretical trends have been, in approximate chronological order, realism, behaviouralism, neorealism, neoliberalism, world systems theory, critical theory and postmodernism. In the years following World War II, political scientists and scholars of international relations increasingly turned to the methods and premises of scientific enquiry. Analysts were more inclined to study recurring patterns in world politics rather than a single event as a focus of enquiry.
- New theoretical models will no doubt be forthcoming in the future.
- It is now widely accepted that, given the range and complexity of the subject matter, a wide variety of theoretical approaches might be regarded as an asset rather than a liability.

Indexing Rules

The above unique features of the disciplines should be taken into consideration when designing the following rules for the indexing of both politics and international relations texts.

Use a Thesaurus

Both subjects rely heavily on a thesaurus for the key elements of using indexing terms. However, flexibility in the form of free language becomes an essential component in the indexing approach, as will be seen later. For example:

> Electoral Legislation could be found under Electoral Acts; Elections Legislation; Legislation
> Legislation and so would have to be standardised by the use of controlled language.

Use a Keyword System

International Relations in particular focuses on relations between states; therefore a largely country-based or region-based keyword system seems the most appropriate approach. A distinction should be made between foreign relations and foreign policy. Whereas foreign relations implies bilateral or multilateral ties, foreign policy should only be indexed under the country which is applying the policy. For example:

> SOUTH AFRICA
> Foreign relations
> (with) Nigeria
> *And:*
> NIGERIA
> Foreign relations
> (with) South Africa
> *But:*
> UNITED STATES
> Foreign policy
> (towards) Israel

However, this rule should not be applied too rigidly, since international issues such as globalisation, nuclear questions and environmental matters increasingly dominate world politics, often requiring main keywords. For example:

GLOBALISATION
Third World stance

Index under Individual Approaches

The variety of theoretical approaches, models and schools of thought should be reflected in indexes focusing on both politics and international relations. Whereas it is possible to combine these texts under one heading, it seems to be more appropriate to index under the individual approaches. For example:

INTERNATIONAL RELATIONS
Theory
And:
POSTMODERNISM

POLITICS
Theory
And:
REALISM

Use New Keywords

Current developments in the political and international arena occur constantly, and unexpected events suddenly dominate national politics and relations between states, and influence their policies. Examples of these would be the events of 11 September 2001 or the assassination of Sheik Ahmed Yassin by Israel. New keywords have to be devised and standardised. For example:

UNITED STATES
World Trade Centre attack
11 September 2001
And:
UNITED STATES
World Trade Centre attack
11 September 2001
impact on Africa

Adopt a Neutral Stance

A neutral stance has to be maintained, differing in this respect from disciplines such as the natural and exact sciences, where objective indexing is easily achieved due to the nature of

the subjects. As social science disciplines, politics and international relations often focus on conflicting ideologies, belligerent actions and violations of the international code of conduct. The indexer should use terminology which is restrained and objective, without compromising accuracy. Two examples follow:

MIDDLE EAST
Conflict
Suicide bombers
Rather than:
MIDDLE EAST
Terrorism
Suicide bombers
or
SOUTH AFRICA
Anti-apartheid organisations
Rather than:
SOUTH AFRICA
Terrorists

Use Cross-references

Cross-references should be liberally applied. Countries, provinces and cities are occasionally renamed, and the titles of regimes, political parties, regional and international organisations are always changing. Acronyms should also be cross-referenced. Multifaceted issues should be approached from various angles, such as the attack on the World Trade Centre. Examples:

BURMA
See also:
MYANMAR

NEW PARTNERSHIP FOR AFRICAN DEVELOPMENT
See:
NEPAD

WORLD TRADE CENTRE
Attack
See:
UNITED STATES
World Trade Centre attack

11 SEPTEMBER 2001
See:
UNITED STATES
World Trade Centre attack

Incorporate Dates

Dates are often the only distinguishing feature between recurring events and should be incorporated. Coup d'états, elections and conferences should be dated. Examples:

MALAWI
Elections 1999
Elections 2002

And:
UNITED NATIONS CONFERENCE ON SUSTAINABLE DEVELOPMENT
Johannesburg, 2002

Indexing Guidelines

- Indexing of theoretical material can be standardised, as these frameworks are developed over time, and even new trends become familiar, for example the trend of postmodernism.
- The same is true of recurring themes – long-term structures result in the developments of a consistent and well-designed indexing terminology, including the appropriate cross-references. The construction of an in-house thesaurus, to ensure consistency, is recommended.
- The indexing of unique and single events appears to pose greater challenges and require care and circumspection. As such developments often result in a string of further events, for example Fall of the Berlin Wall – end of Communism in the Democratic Republic of Germany – a united free Germany. The sequence should be apparent in the construction of an index.
- Both politics and international relations are to a large extent dependent on individual personalities; therefore biographical documentation of participants abound. Information on personalities should be indexed to include distinguishing features which will prevent confusion at a later stage, for example George Bush Snr and George Bush Jnr, both presidents of the United States, should be distinguished by their dates of birth and/or their full second names.

Brief Histories of Two Libraries

Within the Southern African region, two specialised libraries are synonymous with resources on International Relations and Elections. These are the Jan Smuts House Library of the South African Institute of International Affairs (SAIIA) and the EISA Library (EISA was formerly known as the Electoral Institute of Southern Africa), both of which are situated in Johannesburg, South Africa. The former has regional offices in the Democratic Republic of Congo and Mozambique with a third to be opened in Angola.

Jan Smuts House Library

Initially a card catalogue was established (1976), using a country-based keyword system as it was the most practical method to retrieve information. It includes periodical titles, volume and issue numbers, dates and pagination only – neither authors nor titles of the articles were included due to space and time considerations.

Fifteen years later, the Q and A system was introduced as the library's entry into the electronic era. Within a few years, this system had also become obsolete. Since it was necessary to use software with other library functions, such as periodicals cataloguing, archives recording and book ordering, an InMagic package was installed. In 2004, it totalled some 15 000 individual headings and more than 100 000 entries.

The indexing of periodicals in this manner has served the International Relations community and related fields for nearly 30 years. It has ensured effective and specific information retrieval in a world now dominated by the Internet.

The EISA Library

This library was established to develop and sustain an integrated information and research service on election-related issues and democratic transitions in the SADC region. This objective has now been extended to include the entire African continent and beyond, hence the name change.

Indexing was initially nonexistent, as the library consisted mainly of unprocessed material, housed in a series of boxes. An in-house software program was developed, based on experience gained in the Jan Smuts House Library.

Unlike the Jan Smuts House Library, however, it includes a full reference citation, annotation and a multitude of keywords that comply with the specially constructed thesaurus. This ensures precise information retrieval without excessive cross-referencing.

Conclusion

As can be seen from the above, the indexing of material on politics and international relations is an intertwined activity, illustrating both the internal and external elements of the disciplines. In the main, indexing is relatively straightforward; consistency is of paramount importance – as is the ability to adapt to the ever-changing political landscape. Without detailed and meticulous indexing, given the burgeoning literature on the subjects, it would be virtually impossible to present the timeous and precise information needed by scholars, researchers and practitioners.

BIBLIOGRAPHY

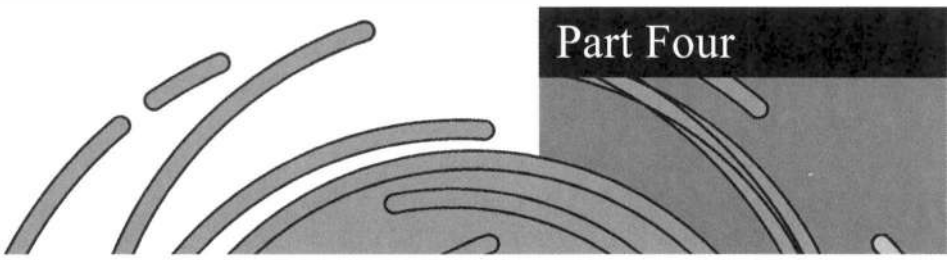
Benyon, J & Jeffery, C. 2000. *Golden times for political studies*. Available: http://www.psa.ac.uk/about/abt gold time.htm (accessed 26 March 2004).

Comfort, N. 1995. *Brewer's politics: a phrase and fable dictionary*. Rev ed. London: Cassell.

Evans, G & Newnham, J. 1998. *The Penguin dictionary of international relations*. Harmondsworth: Penguin.

Knorr, K & Rosenau, JN (eds). 1970. *Contending approaches to international politics.* Princeton: Princeton University Press.

Tanter, R & Ullman, RH (eds). 1972. *Theory and policy in international relations.* Princeton: Princeton University Press.



TRAINING

31 TRAINING OPPORTUNITIES

Marlene Burger

Abstract

Indexers need a good general education, advanced subject knowledge, and specialised acumen around indexes and indexing. Whenever possible, these general educational requirements should be met while still at college, technikon or university and before employment. Indexing skills are taught both at university and by indexing organisations. In general, indexing often forms part of a degree programme in Information Science. Formal education and training of abstractors is seldom offered. In this chapter, education and training opportunities are assessed for indexers and abstractors in Southern Africa. Suggestions and arguments offered by Fourie (2000; 2002) and Theron (2003) are incorporated in this chapter.

Introduction

A career in indexing and abstracting can be a satisfying one, allowing practitioners to develop unique skills and to work in an intellectually challenging environment. These skills can be acquired by formal education, self-education, and by training in the workplace. Formal education may comprise courses or modules on indexing and abstracting offered at universities, or short courses offered by professional associations or individuals. Self-education is accomplished by studying books and manuals in the field.

Training in the workplace is valuable in the sense that very few librarians or information science graduates enter the profession with the explicit aim of becoming an indexer or abstractor. Courses forming part of a degree programme provide limited hands-on experience for students. A great deal of indexing and abstracting experience is needed for full professional development. Short courses, continuing education courses, and workshops give a quick introduction to the fundamentals of indexing and abstracting, followed by a range of practical exercises. Distance teaching programmes can offer working adults valuable opportunities to further their careers in a dynamic indexing and abstracting environment that includes database indexing, book indexing, periodical indexing, web indexing and thesaurus construction.

People who enter the indexing and abstracting field come from many fields, and may be writers, editors, researchers, lecturers, housewives or librarians. Some make a full-time career of it, while others use indexing and abstracting to supplement their incomes. Cleveland and Cleveland (2001:243) list five areas where indexers and abstractors may find work. In the list below these areas are marked with an asterisk (*); the others were added by the author of this chapter:

- *freelance book indexing
- *indexing for independent information brokers, who often use indexers, either on their staff, or on an assignment basis
- *working as library staff members, either indexing special materials in a research library or in a wide range of special libraries
- *working for primary information publishers of scholarly journals, for example
- *working for secondary information publishers, for example in indexing and abstracting services
- working in archives and museums
- working on bibliographic databases and networks, the Internet and the World Wide Web
- working on private large collections in private homes (e.g. Africana, music, photographic items)

Succeeding as a freelance indexer entails more than just being a good indexer. Freelance indexers have to set up a small business enterprise. The advantage of this is the independence of working for oneself. The disadvantage is that there are no guarantees of a steady income and job security. Disciplined work habits and business skills are prerequisites. The most important feature is to promote oneself. A major key to successful freelancing is to be able to sell one's services and convince authors and publishers of the need for indexers. Marketing could aimed at commercial publishers, professional associations, academics, libraries and the like.

The traditional publishing world will continue to need indexers and will remain a source of work, but there are other options – the electronic world of the Internet offers indexing opportunities in hypermedia indexing, electronic publishing and image databases. There is a demand for indexers in contemporary society. De Robbio and Marini (2003:1) have the following to say: '[I]n the ever more pervasive and connected world of digital information, reliable connections among such knowledge representation and lexical, as well as IR, tools such as classifications, lists of subject headings, thesauri, terminological collections and ontologies, are a necessity for networked knowledge-based activities. Users in different settings and with different demands and expectations want to fulfil their information demands wherever information is available, ... , regardless of the heterogeneity of sources'. The more-than 800 000 000 web pages available on the Internet (IMLS 2003) plus all the other published resources and the contents of archives and museums – all this illustrates the enormity of the task awaiting indexers. There is indeed a need for well-educated and trained indexers, as illustrated by the demand for their services. For Southern Africa to become competitive in the international arena, access to information or information-for-action is very important.

The tasks of conceptual analysis and translation of concepts into the indexing language of the system used are based on a large field of theory, for instance the theory of knowledge creation, its existence, storage, retrieval and use (epistemology, ontology, etc.). To be a competent indexer thus requires a sound theoretical schooling. In the present technologically obsessed society, knowledge of technology, entrepreneurial skills, human resources and office management is vital.

Competencies

Various lists of competencies that successful indexers should possess have been compiled by authors of leading indexing manuals. Such lists are not recipes for the education and training of indexers – they are rather wish lists or ideal visions. It is important to note that these competencies could be taught by various education methods such as formal, informal or in-service education and training. This education can be provided by many different types of educators and trainers such as universities, professional societies, and employers. The following is a list of competencies derived from the subject literature and the author's own experience (Theron and Davies list these competencies in chapter 2, but perhaps it is a good idea to repeat them in this the last chapter):

- A good general education in information science, which can also include librarianship, archival science, museum science, publishing and/or any other academic subject(s).
- A sound understanding of knowledge and related concepts such as data and information, for example how information and knowledge are generated, produced and regulated in society. Other matters may involve the recording, organisation for retrieval of information, the search for and use of information in society.
- A thorough understanding of the theory of indexing, its role in providing access to and communication of information; and a thorough understanding of the links between indexing, classification, the assigning of subject headings and thesaurus descriptors.
- Language skills such as semantics and syntax.
- Ability to differentiate between types of indexes such as catalogues, book indexes, databases, manual and automated indexes, and indexes to the Internet. Knowledge of indexes to electronic records is becoming increasingly important.
- Knowledge of information in all forms and formats such as books, pictures, periodicals, museum objects, CD-ROMs, technical reports, newspapers, and posters (i.e. ability to answer the question 'What can be indexed?').
- A proper understanding of the indexing process, that is conceptual analysis and translation. Indexers should be able to apply their knowledge and successfully complete a number of different indexes, for example to books and electronic records. They should also demonstrate an understanding of and ability to use manual and automated (or electronic) indexing methods and systems of indexing.
- Editing and proofreading skills needed for the technical presentation of an index.
- A qualified indexer should be able to undertake user studies, to evaluate indexes and indexing systems, to be able to select appropriate indexing systems or indexes to modify and improve existing ones in accordance with user needs and requirements.
- An understanding of the role and use of technology in indexing, including computers, telecommunications and the requirements of the digital library represents essential abilities for a successful indexer in the digital age.

- The completion of a practical component (especially where formal education is concerned) in the form of an internship, practical sessions at book publishers or Internet providers, et cetera. Ability to compile different kinds of indexes may also be required and can be a valid learning outcome.
- Entrepreneurship, especially for freelance indexers – quoting prices for indexes, marketing their services, knowledge of accounting, budgeting, time and stress management, human resource management, office technology and administration, all comprise valuable skills that should be included in indexing curricula.
- Indexers (including prospective indexers) should also be taught professional and ethical conduct as well as the distinctive characteristics of a profession. These include issues such as service orientation, punctuality, confidentiality, ability to recognise, and acknowledge to clients, that a certain kind of index may be impossible to compile, and ability to handle a clash of interests. Although these issues may be considered peripheral, they are important to the survival of a professional indexer.
- Prospective indexers should also be made aware of the importance of continuing professional education and lifelong learning.

The obvious question now is 'How can all this be achieved?'. How can ideals be turned into reality within the constraints of Southern Africa's education and financial environments?

Short Courses and Workshops (Informal Education and Training)

At present, the Association of Southern African Indexers and Bibliographers (ASAIB) offers workshops in book indexing, periodical indexing (both manual and embedded indexing), and abstracting and reference techniques (bibliography). ASAIB plans to expand these workshops to include thesaurus construction and web indexing. The necessity for these workshops arose from the need for hands-on training in indexing and abstracting that covers a wider field than formal information science teaching programmes.

The indexing workshops are based on the BS ISO 999 (see the bibliography) and include editing, proofreading and guidelines on how to approach freelance indexing. Many indexing manuals have been studied in order to compile the course content (some of the most prominent ones are listed in the bibliography). The workshops are constantly adapted to keep track of developments in the field. ASAIB also maintains a directory of available indexers with their specific subject areas/specialisations indicated. The directory is available on the ASAIB website: www.asaib.org.za

A typical workshop may cover the following topics:

- definitions of core concepts
- characteristics of an indexer/abstractor
- conceptual analysis
- function of an index/abstract

- types of index/abstract
- indexing/abstracting principles
- decisions to make about the planned index/abstract (including policy)
- terminology
- formulating entries
- arrangement and presentation
- copy preparation
- editing
- proofreading
- index/abstract evaluation
- guidelines for freelance indexing/abstracting
- personal relationships

Throughout the workshops, participants work on a variety of exercises. There are also opportunities to evaluate one another's indexes or abstracts. Workshops in embedded indexing include the basic principles of indexing as set out above. However, a good working knowledge of either MSWord or WordPerfect is a prerequisite for enrolment in these workshops. ASAIB presents workshops in Pretoria, Cape Town and Durban in cooperation with the ASAIB Western Cape Branch and the ASAIB KwaZulu-Natal Branch. The workshops are open to any practising or prospective indexer/abstractor. Certificates are issued to participants. Because of the costs involved, ASAIB cannot as yet offer training via the Internet. However, the current Executive Committee of ASAIB is looking into this matter.

The ideal would be for ASAIB to act as the accrediting agency for indexers in Southern Africa. This would involve keeping a register of accredited indexers. Accreditation would require successful completion of an ASAIB training programme, with assessment of practical tasks and a written test. ASAIB may also set the standards for education and training programmes offered by teaching institutions. But this will take a long time to achieve, for the following reasons:

- Need to meet specific requirements regarding South African teaching programmes, such as those of the National Qualifications Framework (NQF), the South African Qualifications Authority (SAQA), the Skills Development Act, the Sectoral Education and Training Authority (SETA) and recognition of prior learning (RPL) requirements. No teaching programme (with a formal qualification as outcome) can be implemented without the approval of these bodies.
- Some programmes (which include indexing and abstracting modules) may already have been approved by these organisations, but the task of ASAIB is then to determine whether those modules comply with the standards envisioned for accreditation purposes.
- Collaboration with teaching organisations, expert indexers, abstractors and bibliographers, the publishing industry and overseas indexing societies will be of paramount importance. ASAIB has to ensure that the standards of its teaching programme compare with those of overseas indexing societies.

- The last survey of training in indexing in South Africa was Pienaar's in 1998. A new, in-depth survey will have to be undertaken, including the whole of Southern Africa.
- Employers of indexers, abstractors and bibliographers in Southern Africa will need to recognise ASAIB's accreditation. This will call for an extensive marketing campaign.
- An option is to franchise the training programme of the Society of Indexers for example, but there are problems such as its British bias and the time and effort that would be involved in adapting the programme for Southern African purposes, as well as the cost.

ASAIB is considering all the possible options for achieving accreditation.

Distance Education (Formal or Informal)

Indexing and abstracting are both complex processes involving higher order cognitive skills. They require considerable practice, considerable reflection, and for learner indexers/abstractors to be critical of the quality of their own work. Not only are learner indexers/abstractors required to apply the guidelines and rules as explained in standards and textbooks, but they are also required to perform conceptual analysis to identify key concepts. This is a difficult process that can leave the novice indexer/abstractor feeling some uncertainty. It is also important for indexers and abstractors to be aware of diverse information needs and to react accordingly. Indexing and abstracting are furthermore subjective rather than objective processes, which means a single individual may, for example, assign different indexing terms at different times to the same text, or more than one individual may assign different terms to the same text. Although both can be conducted as paper-based exercises, they can also be done electronically.

The teaching of indexing and abstracting therefore poses a number of challenges, especially as they are also marked by new developments such as changes in indexing software, and the demands of a variety of indexing and abstracting types (e.g. web indexing, critical and structured abstracts). Although many standards and textbooks (e.g. Booth 2001; Borko & Bernier 1978; Cleveland & Cleveland 2001; Lancaster 1998; Mulvany 1994) offer advice on the steps and principles to consider for both indexing and abstracting, the main problem lies in the application of these skills, for example having to decide what an entity is about and representing it in vocabulary that will meet users' needs and support IR.

Apart from practical skills there is also a need for more intensive research in indexing and abstracting. Anderson (1994) remarks: '[T]here is so little research and so much ignorance about indexing, especially on how people use indexes, and what constitutes good and useful indexing', while Millstead (1994) declares: 'People have been making indexes for hundreds of years, but there has been relatively little research on indexing as such.' Since 1994 (when these sources were compiled) the scene has changed somewhat, with the focus rather on organisation of multimedia, visual presentations of information organisation, use of thesauri, natural language and information organisation in digital libraries, rather than on cognitive (thinking) skills.

The essence of distance education lies in the geographic separation between trainer and learner, and the fact that they are not in the same place at the same time. Learners are

seldom in the position where they can ask questions and get immediate, direct answers from the trainer, except when video conferencing or web-based bulletin boards are used.

Distance teaching has both advantages and disadvantages for the training of indexers and abstractors. The following are some of the advantages: trainees can carry on with their jobs and family responsibilities; trainers and trainees can stay in remote areas; geographical distance does not keep people from furthering their careers; trainees can study at their own pace; trainees can study at times which they find convenient (the same applies to the trainers developing the course content); fees are normally cheaper and there are no, or few, travelling expenses; one can use expertise from all over the world to help develop course content and to do assessment; a variety of teaching methods can be used, selected and combined according to the needs of the trainees; one can cater for individual differences; training can be offered to people who otherwise could not afford formal training or attend workshops.

Disadvantages include the following: it is not easy to teach basic technical skills (e.g. using a computer); trainees must have good reading and writing skills, and a strong sense of self-discipline and responsibility; sometimes restrictions are imposed, for example dates for tests (however in some distance teaching programmes learners can be tested whenever they are ready and write tests online); one needs to assume all possible scenarios for the learners, their experiences and potential questions (probably without ever meeting them); there is seldom direct synchronous communication between learners and trainers (learners may feel isolated); they must not only have access to the appropriate IT infrastructures, but must also be familiar with the IT that will be used; developing and maintaining distance teaching material is very time-intensive and labour-intensive. There should also be opportunities for learners to assess their own progress, answers for self-assessment exercises and other activities should be provided, with a wide variety of examples from the real (indexing and abstracting) world.

Instructional Design

The planning of a distance teaching programme is influenced by three key needs:

- need to understand the concepts of indexing and abstracting, the practice and theory of these concepts, and developments and research findings
- need to interpret the changing indexing and abstracting environment and to develop lifelong skills
- need to understand and interpret the concept of distance teaching, and the potential value of distance teaching for the training of indexers and abstractors

The way training is planned is also influenced by the view on teaching and training, and of the people trained. This is one reason why there are so many different approaches to training indexers. Philosophy of teaching can also be called the epistemology of teaching. Is training merely the provision of training materials, or does it include offering examples and practical exercises, and does it include feedback and assessment? Furthermore, does it perhaps include extensive support and enrichment?

Instructional design should be the point of departure when considering the possibilities for the distance teaching of indexers and abstractors. The key principles of instructional design also apply to the design of practical workshops for indexing and abstracting. When making the final decisions on how to deliver the programme (e.g. through print-based training material or through the WWW), one should also consider the characteristics of distance teaching and its specific requirements and challenges.

Instructional design can be described as a logical, step-by-step approach to the planning of training programmes. It helps one consider everything that may have a bearing on the training. Seels (1995:xi-xii) says: '[It is] the process of specifying conditions of learning. Its purpose is to create strategies and products on the macro level, such as programs and curricula, and at the micro level, such as lessons and modules.' It is most often guided by a model of instructional design – many models can be found in the subject literature (e.g. Dick & Carey 1996; Kemp, Morrison & Ross 1998). These models can be adapted according to one's personal preferences and circumstances. The model depicted below is an adaptation of a model presented by Ina Fourie in 1995:

An instructional design model will help one to determine *inter alia*

- for whom the training programme will be developed (e.g. Library and Information Science [LIS] students, novices or experienced indexers and abstractors)
- what one wants the learners to learn or demonstrate, that is what they should be able to do after completing the training programme
- the best way to learn the subject content or skill
- the best way to offer the programme (e.g. through print or online)
- the extent to which the learning has been achieved (the assessment procedure)

Steps	Questions to answer
Situation analysis	Why is training necessary? What is the indexing and abstracting environment like? Who will be trained? Who will offer the training? What training materials are available? What technology is available? Which factors may impede the training?
Formulation of the aim of the training and the learning outcomes	What will be the aim of the programme? What should the training cover? What will be the learning outcomes, i.e. what should the learners be able to do after completing the training? What should their entry knowledge be? How will the learners be assessed?
Designing of the training material	Which teaching approach will be best? Which teaching methods will be best? Which teaching strategies will work best? Which media will be used to deliver the programme? How will you communicate with the learners? How will you allow for individual preferences and circumstances? How will you allow for the learner's contribution? How will you develop a supportive and stimulating climate?

Steps	Questions to answer
Development of training material	This includes the preparation, duplication and distribution of the material (e.g. workbooks, videos, websites).
Training	This includes marking assignments or portfolios, offering feedback, presenting classes and workshops, answering questions and developing personal rapport with the learners.
Evaluation	This includes the evaluation or assessment of the learners' progress, as well as the evaluation of the success of the training programme.

Situation Analysis

Within the context of a particular environment, a situation analysis should shed light on the decisions to make in the planning of the programme. The indexing and abstracting environments are characterised by change – indexing methods, for example, have changed. The indexing process and therefore also the training of indexers should always be seen as part of the bigger picture. There have also been calls for closer links between the theory and practice of indexing and abstracting and the overall indexing process. Bates (1995), for example, declares: 'Finally indexing and information system design need to proceed within the context of a deep understanding of the character of the intellectual domain, its culture, and its research questions.' Unfortunately there is not always sufficient time to do a formal situation analysis. Even if one has to rely mostly on one's gut feeling and personal perceptions, it is still worthwhile to at least consider the issues forming part of the situation analysis (see table above). The information collected during a situation analysis forms the basis for subsequent decisions. These decisions will also be influenced by one's views and convictions in this regard.

What to Teach

In addition to the components of the situation analysis, it is also necessary to identify learners' needs. This is called a needs assessment. Needs are the gap between what is expected and the existing conditions. In other words, they reflect the gap between the current situation and where one would like learners to be after completing the training programme. The following are examples of needs: indexers requiring training in the use of indexing or thesaurus software, or on how to start an indexing or abstracting business, how to evaluate the effectiveness of an index or database, how to interpret and apply research results in their indexing and abstracting practices, how to use the Internet to keep track of indexing and abstracting developments, or how to analyse a document. A needs assessment can be done by means of questionnaires, interviews, brainstorming sessions or focus group interviews. Indexing and abstracting standards and textbooks may also offer guidelines. In order to offer dynamic teaching programmes, the processes of indexing and abstracting, specifically the tasks of an indexer/abstractor, should be reassessed.

Task Analysis

The purpose of a task analysis should be to identify one or more of the following indexing and abstracting matters:

- tasks
- methods employed
- cognitive (thinking) skills of experts
- tools used
- order of activities that need to be completed
- relationship between different activities (e.g. conceptual analysis and the selection of terms)
- mental model (framework)

The following are task analysis methods: task descriptions, methods analysis, learning hierarchy (prerequisite analysis), information processing analysis, learning contingency analysis, activity theory and conceptual graph analysis. For more details on these methods, Jonassen, Tessmer and Hannum (1999) can be consulted. Doing a task analysis will help clarify the complexities of indexing and abstracting. The most difficult part of indexing is not applying rules, such as arranging the entries in alphabetical order, constructing cross-references, or using the plural form for countable nouns: 'Deciding what an entity is about, and how to select the most appropriate indexing terms, are the problematic issues that we still do not know how to teach. How do you teach the skills of conceptual analysis? I believe that task analysis can shed light on this, and it will probably point us in new directions when selecting teaching methods and teaching strategies' (Fourie 2000). The information gained from task analysis can also help one take decisions (e.g. on the theory behind indexing) such as deciding between mission-oriented indexing, request-oriented indexing, subject-oriented indexing and entity-oriented indexing. Is it necessary to offer training in such a way that the learners can experiment with different approaches, learn from their mistakes, and gradually improve the quality of their work?

Teaching Aims and Learning Outcomes

Examples of teaching aims could include the following: to train learners to prepare indexes or abstracts that will support effective information retrieval, to contribute to the knowledge of information retrieval, to keep track of future developments, to analyse specific web-based information systems, to determine the need for web-based indexing, or to decide which vocabularies to use. Learning outcomes should be formulated clearly – it is extremely important for distance learners to know *exactly* where they are heading and what is expected of them.

Different taxonomies can be used to formulate learning outcomes, for example Bloom's taxonomy which is well known and always a safe choice. He distinguishes between the cognitive, psychomotor and affective domains. Knowledge of something represents the most basic level, while evaluation represents the most difficult level of the cognitive domain (Kemp, Morrison & Ross 1998). Examples of applying Bloom's taxonomy include the following: after completing the training the learner should be able to explain the difference

between derived indexing and assigned indexing, analyse a selection of entity types (e.g. a periodical article, a report), assign appropriate indexing terms and write an informative abstract, and critically compare the use of natural language versus controlled vocabulary. When considering the changing indexing environment and the demands made on indexers, it appears that we need a new taxonomy to deal with these challenges. Jonassen and Tessmer have suggested a new taxonomy that seems to have potential for the training of indexers. For details of this taxonomy, consult the book by Jonassen, Tessmer and Hannum (1999).

The following are examples of learning outcomes: after completing the programme, learners should be able to explain core indexing terms and their relationship, evaluate the effect of indexing on an information retrieval system, the impact of their mental models of an information retrieval system on their indexing practice, devise an indexing policy for a specific real-life indexing case study or a proposal for their training needs. As the learners may be adults and experienced indexers and abstractors, one could also allow them to add their own learning outcomes.

Training Programme

In the past, mainly behaviouristic approaches have been used to set teaching objectives. Once learners have mastered an objective, they move on to the next objective. In the dynamic indexing environment, constructivism seems to offer a more viable approach. Here learners are given opportunities to build their own knowledge through personal interpretation. The skills of reflectivity and not just memory or memorisation are emphasised. Constructivism also encourages learners to be more critical about their own work. Although this approach can be used for print-based distance teaching, it is much more effective with online methods where the trainer communicates with learners and learners can communicate with each other. Constructivism seems to offer an ideal teaching approach to the training of indexers and abstractors, since indexing and abstracting can never be perfect. There are also the factors of inter-indexer and intra-indexer inconsistency to consider. Users' needs are also constantly changing. Ongoing self-reflection and improvement are crucial to indexing and abstracting. The constructivist approach allows for such improvement without penalising the learner. However, it does not imply that the final product (e.g. an index or abstract) should be of inferior quality. It is only the *process* of getting there that is made more supportive and more tolerant of the learner's efforts to acquire new knowledge.

Distance teaching can be based on one method (e.g. print-based study guides or manuals) or a combination of methods. The material can be delivered in print-based or electronic format (e.g. web-based instruction or CAI). According to Fourie (2000), the following methods and materials are all suitable for distance teaching:

- Study guides/manuals contain the text with explanations, theoretical background, different viewpoints, and so on, as are normally provided in a lecture.
- Workbooks contain examples, practical exercises and solutions.
- Textbooks offer a wider perspective on problems of and different approaches to indexing/abstracting.

- Videos offer detail on information retrieval (i.e. information searching) and how this can be supported by indexing/abstracting.
- Computer-assisted instruction (CAI) can be used to illustrate the effect of indexing/abstracting on information retrieval and on different points of view on, say, conceptual analysis.
- Discussions can be used to compare different views of indexing/abstracting.
- Case studies may involve working with existing databases, evaluating the indexing/abstracting of a website, or evaluating book indexes or abstracts of periodical articles.
- Expert guest trainers can be used to share different viewpoints (e.g. through a website).
- Mentors can be used to offer on-the-job advice and support.
- Exemplars of databases, indexes and websites used in practice can offer a view on real-life problems.
- Reviews of indexes, for example book indexes, can be used to help learners to refine their indexing skills, and to learn from experts.
- Networking (i.e. interaction with other learners through study groups or web discussions) can be used to stimulate thinking and to collaborate with other learners. It can also be used for peer assessment.
- Supportive tools, for example access to demonstration versions of database or indexing software, online reference works and online thesauri can enrich the learning experience.
- Enrichment can also be offered through links to resources such as websites of indexing societies, e-journals, electronic discussion groups, and publishers and database producers who employ indexers or abstractors.

Most of these methods and materials can also be used in print-based approaches to distance teaching. The most exciting possibilities are with web-based instruction where the study material and learning opportunities can all be available interactively. The choice of an appropriate method or combination of methods will depend on the results of the situation analysis and on the task analysis. One should always use a well-balanced mixture of media.

After deciding on the method or combination of methods to use, a decision must be made on appropriate teaching strategies that will help learners achieve the learning outcomes. A teaching strategy prescribes the sequence of events and offers guidelines on when and how the content should be presented to achieve the learning outcomes. Among other things one has to select strategies that allow for the distance between trainer and learners, individual needs and preferences, support, two-way communication, the contribution adults can make to the learning situation, pre-instructional strategies (e.g. initial motivation, explanation of the aim and learning outcomes, overviews, specification of the entry or prior knowledge level), information presentation, learner participation, et cetera. The teaching strategy should also make allowance for assessment, follow-through activities, collaborative learning, working with real-life, poorly structured problems (e.g. case studies), opportunities for learners to construct their own meaning and to experiment with different solutions and points of view (e.g. by using portfolio assessment and journal writing), collecting information on the learners' experiences, and opportunities to practise and get feedback. Although teaching

strategies should be carefully planned, one should also take care not to go into so much detail that there is no time left to write and construct the actual study material.

Different assessment methods can be used in the training of indexers and abstractors. These include a variety of self-assessment questions and exercises, practical indexing/abstracting exercises and formal tests.

The last steps concern formative and summative evaluations. Formative evaluation is extremely important during the planning stage. One can, for example, get feedback from colleagues and expert indexers/abstractors, as well as from potential learners in the programme. Summative evaluation considers the overall effectiveness of the programme in helping learners achieve the learning outcomes. If necessary, the programme should be adapted or revised.

If distance teaching is preferred, indexers in Southern Africa have to enrol for courses offered by overseas indexing societies such as the British Society of Indexers (SI). Their Distance-Learning Course in Indexing is available on CD-ROM and consists of five units. For enquiries, e-mail can be send to admin@indexers.org.uk or you can access their website at www.indexers.org.uk Other indexing societies to contact are

- American Society of Indexers (ASI) at info@asindexing.org or www.asindexing.org
- Australian and New Zealand Society of Indexers (ANZSI) at memsec@aussi.org or www.aussi.org
- Indexing and Abstracting Society of Canada (IASC) at www.indexingsociety.ca

Proposed Models for a Training Programme

Theron (2003) suggests two possible models for the training of indexers and abstractors in Southern Africa. However, some issues must be kept in mind before analysing these models: the demographic profile of potential learners (gender, race, place of residence, financial resources and existing educational levels); the requirements of the NQF, SAQA, SETA and RPL; the assumption that there is a need for professional indexers; and what level of education is required. A workshop or brief hands-on course may be sufficient for an experienced indexer/abstractor, especially when it is used as a form of continuous professional education. For a first-level student who has just matriculated, a completely different education model is needed. Such learners may be disadvantaged as far as language skills, general knowledge, reading and comprehension are concerned. Against this background the design of an education model in order to reach the core competencies (as set out earlier in this chapter) becomes quite complicated.

In this model, library schools and information science departments take responsibility for the education and training of indexers and abstractors. Indexing and abstracting form part of the curricula for the professional education of information professionals such as librarians. Often indexing and abstracting is only one section of a major such as subject organisation (including classification, assigning of subject headings, thesaurus construction, etc.) – it is seldom offered as a major on its own. This immediately implies that only the basics are taught with limited practical experience. Even when a practical internship is required from information science students, they rarely get the opportunity to do actual indexing.

The challenge facing library/information science departments is to incorporate more opportunities for practical work (i.e. compiling and evaluating indexes and abstracts). A solution might be Fourie's (2002) socio-cognitive approach (for details read the article listed in the bibliography). The most viable solution seems to lie in brief hands-on courses and workshops supplementing the formal education programmes of teaching institutions. Specialisations can be offered, such as Internet indexing, use of dedicated indexing software such as Macrex or Cindex, embedded indexing, and so on. Employers (e.g. libraries, publishers) can also consider presenting in-service training for employees as part of a staff development programme. ASAIB offers workshops supplementing formal educational programmes for indexers and abstractors.

The second model, the Information Architecture Model, provides an interesting perspective that needs further investigation. Belton (2003:1) describes it as follows: 'Information architecture is currently emerging as a discipline that concerns itself with the development of systematic approaches to the organisation and presentation of on-line information.' This approach places the education and training of indexers and abstractors among that of the other design professions. Most information scientists tend to think of information as content only, and not as something that has a form or architecture. This form, organisation and structure of information comprise the substance of information architecture.

This model has important pedagogical implications for the education and training of indexers and abstractors – it implies that professional education supports the three dimensions of professional attributes, that is the cognitive, the normative and the evaluative. The cognitive dimension is centred on the body of knowledge and techniques needed by a professional to apply his or her work as well as the study and training necessary to master this knowledge; the normative dimension refers to the service orientation of the professional indexer/abstractor and ethical approaches to his or her work; and in the evaluative dimension the differences between a professional's occupation, other types of work or professions and his or her own are constantly compared to emphasise his or her profession's uniqueness, autonomy and prestige. For more information about this model, read the article by Belton listed in the bibliography.

The following are the implications of this model for the education and training of indexers and abstractors:

- Indexers/abstractors must be taught the body of knowledge unique to their profession (e.g. information infrastructure, production and use of information, indexing and abstracting as part of the process, the information environment, research, theory building).
- More emphasis should be placed on the human aspects of indexing/abstracting. Indexes and abstracts are compiled for humans and for use by them. Service orientation and ethics form part of this knowledge.
- A clear conceptual and theoretical grasp of the essence of the profession of the indexer or abstractor is essential. This ensures that indexers or abstractors do not trespass on other professions' domains, and justifies the existence of the indexing/abstracting profession in society.

For training purposes this means that prospective indexers and abstractors need to be exposed to proper theoretical education. This education should include the theory and philosophy of education, information, correct professional conduct, and the nature of the profession.

Conclusion

In conclusion it can be stated that the education and training of indexers and abstractors is important, mainly because of their central role in providing access to information. Given the dearth of facilities, this is particularly significant to Southern Africa.

BIBLIOGRAPHY

Anderson, JD. 1994. Standards for indexing: revising the American National Standards Guidelines Z39.4. *Journal of the American Society for Information Science*, 45(8):628-636.

Bates, MJ. 1995. Indexing and access for digital libraries and the Internet: human, database, and domain factors. *Journal of the American Society for Information Science*, 49(3):628-636.

Belton, BK. 2003. A design foundation for information architecture. Available: http://eprints.rclis.org

Booth, PF. 2001. *Indexing: the manual of good practice*. München: Saur.

Borko, H & Bernier, CL. 1978. *Indexing concepts and methods*. New York: Academic Press.

Bridgeway Career Development. 2003. *Career profile: Indexer. Interview by Janet Scarborough with Carolyne G. Weaver.* Available: www.bridgewaycareer.com/profile-indexer.htmBritish Standards Institution. 1996. *Information and documentation – guidelines for the content, organization and presentation of indexes.* BS ISO 999:1996. London: British Standards Institution.

Cleveland, DB & Cleveland, AD. 1990. *Introduction to indexing and abstracting*. 2nd ed. Englewood, Colo: Libraries Unlimited.

Cleveland, DB & Cleveland, AD. 2001. *Introduction to indexing and abstracting*. 3rd ed. Englewood, Colo: Libraries Unlimited.

Dick, W & Carey L. 1996. *The systematic design of instruction*. 4th ed. New York: Longman.

Fourie, I. 2000. Teaching indexers and abstractors through distance teaching programmes. Adaptation of a Workshop Offered at the *33rd Annual Conference of the American Society of Indexers, 1-3 June, Boston, titled: Distance teaching of indexers: keeping track of new challenges as well as traditional practices.* (Unpublished version used for this chapter).

Fourie, I. 2002. How can we take a socio-cognitive approach in teaching indexing and abstracting? *The Indexer*, 23(2):83-85.

Fourie, I. 2003. Using distance teaching methods to further careers of working adults. *Behavioural & Social Sciences Librarian*, 21(2):29-48.

IMLS: A closer look: highlights of the month: Archives. 2003. Available: www.imls.gov/scripts/text.cgi?/closer/archive/hlt_11000.htm

Jonassen, DH, Tessmer, M & Hannum, WH. 1999. *Task analysis methods for instructional design.* Mahwah, NJ: Lawrence Erlbaum.

Keegan, D. 2000. *Distance training: taking stock at a time of change*. London: Routledge.

Kemp, JE, Morrison, GR & Ross, SM. 1998. *Designing effective instruction*. 2nd ed. Upper Saddle River: Prentice-Hall.

Lancaster, FW. 1998. *Indexing and abstracting in theory and practice.* 2nd ed. London: Library Association.

Managing cataloguing and the organization of information: philosophies, practices and challenges at the onset of the 21st century. Edited by RC Carter. New York: Haworth.

Manley, S & Harwood, N. 1997. Bringing it home: learning to index books by correspondence. *The Indexer*, 20(4):185-187.

Millstead, JL. 1994. Needs for research in indexing. *Journal of the American Society for Information Science*, 45(8):577-582.

Mulvany, NC. 1994. *Indexing books.* Chicago: University of Chicago Press.

Pienaar, B. 1998. *Training in indexing and bibliography*. Johannesburg: Association of Southern African Indexers and Bibliographers.

Seels, BB. 1995. *Instructional design fundamentals: a reconsideration.* Englewood, NJ: Educational Technology Publications.

Taylor, AG. 2004. *The organization of information.* 2nd ed. Westport, Conn: Libraries Unlimited.

Theron, JC. 2003. The need for the professional education of indexers in South Africa, in *Signposts to knowledge: indexing in the 21st century: papers read at an ASAIB Conference held in Cape Town, 14-15 April 2003.* Edited by E Schoeman. Johannesburg: Association of Southern African Indexers and Bibliographers:33-43.

Wallis, E. 1997. Indexing as a professional activity. *The Indexer*, 20(4):189-191.

LIST OF CONTRIBUTORS

Abbott, Kate
William Cullen Library, P/B X1,
PO Wits 2050
Tel: +27 11 717 1940
E-mail: Katea@library.wits.ac.za

Ambrose, David
National University of Lesotho,
PO Roma 180, Lesotho
Tel: + 266 22213009

Archibald, Carol
William Cullen Library, P/B X1,
PO Wits 2050
Tel: +27 11 717 1940
E-mail: Carola@library.wits.ac.za

Arnold, Anna-Marie
Environtek, CSIR,
PO Box 395, Pretoria 0001
Tel: +27 12 841 2683
E-mail: aarnold@csir.co.za

Burger, Marlene
Department of Information Science,
PO Box 392, Unisa 0003
Tel: +27 12 429 6585
E-mail: burgem@unisa.ac.za

Davis, Gavin R.
Department of Library and Information
Science, University of the Western Cape,
Private Bag X17, Bellville 7535
Tel: +27 21 959 3546
E-mail: gdavis@uwc.ac.za

Du Preez, Madely
Department of Information Science,
PO Box 392, Unisa 0003
Tel: +27 12 429 6792
E-mail: preezm@unisa.ac.za

Duncan, Peter
William Cullen Library, P/B X1,
PO Wits 2050
Tel: +27 11 7171947
E-mail: peterd@library.wits.ac.za

Ferreira, Nico
Unisa Law Library,
PO Box 392, Unisa 0003
Tel: +27 12 429 3352
E-mail: ferrenm@unisa.ac.za

Fourie, Ina
Department of Information Science,
University of Pretoria, Lynnwood Road,
Pretoria 0002
Tel: +27 12 420 5216
E-mail: fouriei@up.ac.za

Garson, Yvonne
Tel: +27 11 646 0700
E-mail: garsonn@social.wits.ac.za

Jagarnath, Omesh
University of KwaZulu-Natal,
EGM Library, King George V Avenue,
Durban 4001
Tel: +27 31 260 2063
E-mail: Jagarnat@ukzn.ac.za

Kalley, Jacqueline A
Electoral Institute of Southern Africa (EISA)
Tel: +27 11 482 5495
E-mail: jackie@eisa.org.za

Kotze, Antoinette
National Library of South Africa
Tel: +27 12 321 8931
E-mail: antoinette.kotze@nlsa.ac.za

Lastovica, Ethleen
31 Silverlea Road, Wynberg,
Cape Town 7800
Tel: +27 21
E-mail: lastoaj@mweb.co.za

Le Roux, Sophia
Sophos Information Services,
PO Box 695, Swellendam 6740
Tel: +27 28 514 3278
E-mail: sophos@intekom.co.za

McGuirk, Karin
Department of Information Science,
PO Box 392, Unisa 0003
Tel: +27 12 429 6313
E-mail: mcguik@unisa.ac.za

Merrett, Christopher
Tel: +27 33 260 5111
E-mail: merrett@nu.ac.za

Musiker, Reuben
SA Jewish Board of Deputees,
PO Box 87557, Houghton 2041
Tel: +27 11 888 4936
E-mail: nmusiker@netactive.co.za

Pickover, Michelle
William Cullen Library, P/B X1,
PO Wits 2050
Tel: +27 11 717 1940
E-mail: Michelp@library.wits.ac.za

Pretorius, Martha
22 Clyde Street, Murrayfield,
Pretoria 0184
Tel: +27 12 803 2618
E-mail: mpretorius24@mweb.co.za

Reddy, Umashanie
Emigrated to Canada in 2005
(no contact details available)
Was employed at University of KwaZulu-Natal, EGM Library, Durban

Robertson, Elizabeth
PO Box 47318, Parklands 2121
Tel: +27 11 484 2620
E-mail: erob1@mweb.co.za

Schoeman, Elna
PO Box 666, Randburg 2125
Tel: +27 11 787 6146
E-mail: elna@internext.co.za

Shier, Judith
322 South Ridge Road, Durban 4001
Tel: +27 31 812 721
E-mail: jshier@worldonline.co.za

Schulze, Malene
National Cultural History Museum,
PO Box 28088, Sunnyside 0132
Tel: +27 12 324 6082
E-mail: malene@nfi.org.za

Theron, Christie
Passed away on 2 January 2005
Was employed at the Department of Information Science, University of South Africa, Pretoria

Underwood, Peter
Centre for Information Literacy,
University of Cape Town, P/B
Rondebosch 7701
Tel: +27 21 650 3091
E-mail: pgunderwood@ched.uct.ac.za

Van der Walt, Thomas
Department of Information Science,
PO Box 392, Unisa 0003
Tel: +27 12 429 6520
E-mail: vdwaltb@unisa.ac.za

Watkins, Ansie
Unisa Library, PO Box 382, Unisa 0003
Tel: +27 12 429 3426
E-mail: watkiapj@unisa.ac.za

Willet, Shelagh
PO Box 20166, Gaborone, Botswana
Tel: +267 355 5156
E-mail: willet@mopipi.ub.bw

INDEX

INDEX

Note: Due to a page restriction, a broad, simple index was compiled, mainly based on the headings in the text. In most instances, only names and titles that appear in headings are included in the index. Some entries are combined, for example *abbreviations/acronyms.* It is arranged in word-by-word order, for example *free text* is filed before *freelance indexers.* Prepositions in subheadings are ignored in the alphabetical arrangement. *See* references are used to guide the reader. Since this manual is in essence about indexing, the subheading *indexing* does not appear under all headings. It was also not possible to use subheadings for all headings with several page references.

abbreviations/acronyms 174, 224, 230-231, 240
aboutness 20, 21, 33, 112
abstracting/abstracts 141-158
  accuracy 153
  automatic 154-156
  bibliographic descriptions 149-150
  comprehensiveness 154
  critical 143-144
  definition 141
  guidelines 149-154
  indicative 141-142
  informative 142-143
  modular 145-147
  omissions 154
  organisation policy 44-49
  reasons for 148
  signing 154
  steps 148-149
  structured 144-145
  style 150-153
  training 307-322
  verbal subject description 57
  words from 106
accuracy
  abstracts 153
administrative metadata 131
African Studies 208
Africana 197-209
  archival records 200
  authority files 208
  books 198-199
  Cape Province 200, 201
  codex records 198-199
  collections 200-201
  definitions 197-198
  ephemera 199-200
  films 199
  Free State 201
  Gauteng 201
  handwriting 202
  illustrations 199
  indexes 201-202
  indexing problems 202-204
  indexing resources 201-202
  KwaZulu-Natal 201
  language 202, 204
  manuscripts 200
  metacodex records 199
  names for events 203
  orthography 203-204

overview 197-198
periods 198
personal names 203
photographs 199
place names 203
reference sources 202
South Africanisms 204
travelers 199
types 198-200
user guide 204
weights/measures/monetary systems 204
alphabetical arrangement
Khoe-San languages 246-247
letter-by-letter 80, 185-186, 259
word-by-word 80, 185-186, 259
alternative press
labour 240-241
analytical chemistry 228-229
antique maps 210-215
care of 211
cartographers 213
cataloguing 212-214
evaluation 211-212
management 210-211
publication 214
antonyms 81
Approved South African Journals 101
archival records
Africana 200
archival science 12
archives 216-220
cataloguing 216
correspondence 220
EAD 217
finding aids 134, 216-218
Historical Papers 218
indexes 216-217
indexing problems 219-220
indexing protocols 218
overview 216-218
photographs 218-219
SGML 217
standards 217-218
TEI 217
thesaurus 217
William Cullen Library (University of the Witwatersrand) 218
WWW 217
arrangement 79-80, 191-192
alphabetical 185-186
Khoe-San languages 246-247
letter-by-letter 80, 185-186, 259
word-by-word 80, 185-186, 259
ASAIB 6, 42
short courses/workshops 310-312
assigned indexing terms 114
Association of Southern African Indexers and Bibliographers *see* ASAIB
audience
labour 239
authoring tools
web indexing 119
authority files
Africana indexing 208
periodicals indexing 106-107
authors' keywords
periodicals indexing 107
automatic
abstracting 154-156
indexing 16, 35, 111-116

Belkin, N 10-11
bibliographic descriptions
abstracts 149-150
bibliographic services 252-254
bibliographies
legal 251-252
national 285-291
biological sciences 230-231
abbreviations/acronyms 230
geographical names 231
scientific/popular names 230
species/family names 230
book indexing 68-85
alphabetical arrangement 185-186
antonyms 81
arranging entries 79-80
components 71-72
cross-references 79, 82, 183
definitions of concepts 68-70
editing 82, 180-182
embedded indexing 86-91
entries 76-80
errors 182
exhaustivity 80-81
final index 82-83
guidelines 73-74
health sciences 234
homographs 81
how to index a book 72
information needed 72-73
interpretation 76-77

introductory notes 72
layout 186
letter-by-letter arrangement 185-186
locators 78-79, 183-184
main headings 77-78, 81-82, 184-185
multi-word headings 182-183
physical entity 75-76
policy 80-82
presentation 180-182
qualifiers 184
reasons for 70-71
recording entries 78
relationships between entries 78
scattering 182
singular/plural 82
software 83-84
spaces 187
specific tasks 182-185
specificity 80-81
standards 73-74
steps 74
studying the physical entity 75-76
style 76-77
subheadings 77-78, 81-82
synonyms 81
technical presentation 179-188
translating the content 76-80
typestyle 187
what to index 74-75
word-by-word arrangement 185-186
books
Africana 198-199
indexes to 5-6
indexing 68-85
literature for children 257-267
Boolean operators 59-60, 112
boundaries
web indexing 119
bundling
automatic indexing 114
Business Blue Book of South Africa 87-91

Cape Town English Press Index 5
cartographers 213
cataloguing
Africana 206
antique maps 212-214
archives 216
museums 268-269
CD-ROMs 4, 87, 88, 90, 97, 208, 214
characteristics
indexes 30-35, 258-261
thesauri 160-163
chemical names 229
children's literature *see* literature for children
CIDOC Guidelines 270
classification/coding
periodicals indexing 107
classification notations 54-57
classification systems 54-55
museum objects 277-278
click sounds
Khoe-San languages 245-246
client services
Unisa Law Library 254-255
codex records199
collection-level metadata 130
collections
Africana 200-201
antique maps 210-211
archives 218
health sciences 234
competencies
indexers 14-16, 220, 292-293, 309-310
components
book indexes 71-72
compound terms 174-175
comprehensiveness
abstracts 154
computer(s)
Khoe-San languages 245-246
museums 269
national bibliographies 288
science 13
systems 111-116
concepts 22
book indexing 68-70
environmental issues 227
expression 34
indexing 130
metadata 130
conceptual
analysis 9, 20-28
approach 88
conceptualisation 9, 20-28
aboutness 21-22
complexity 27
concepts 22
content-oriented approach 24
entity representation 21
facets 26

guidelines 22-23
overview 20-21
requirements-oriented approach 24-25
simplistic approach 24
socio-cognitive view 25-26
views on 23-27
content-bearing words 33
content metadata 131
content-oriented approach
conceptualisation 24
context
indexing 7-19
controlled vocabulary 64, 66-67
words from 106
correspondence
archives 220
critical abstracts 143-144
cross-references 34, 70, 79, 82, 183
literature for children 260
newspapers 295
politics and international relations 301
web indexing 123-124
crosswalks
metadata 135

data standards
MDA 270-272
museum objects 270-271
dates
museum objects 276-277
politics and international relations 302
definitions
book indexing concepts 68-70
embedded indexing 86-87
index 29-30
indexing 7-8
metadata 129
thesaurus 160
derived indexing terms
automatic indexing 115
description meta-tags
web indexing 127
descriptive metadata 131
display
web indexing 119
distance education
training 312-319
document chunking
web indexing 122
Dublin Core metadata 130, 135-139
advantages 135-136
element set 136
template 137-139

EAD 134, 217
editing 82, 180-182
alphabetical arrangement 185-186
cross-references 183
errors 182
factors influencing 181-182
headings 184-185
locators 183-184
multi-word headings 182-183
qualifiers 184
scattering 182
tasks 180, 182-185
education 221-225
acronyms 224
indexing 223-224
indexing categories 222
intellectual skills of indexers 223
South Africa 221-222
terms 223, 224
effectiveness
automatic indexing 113, 114
EISA Library 303
elections 298
electronic indexes 6
hyperlinking 90-91
procedure 89
Quark Index Xtention 89-90
embedded indexing 86-91
Business Blue Book of South Africa 87
CD-ROM 90
conceptual approach 88
decisions 89
definition 86-87
hyperlinking 90-91
locators 118
preparations for 88
procedures 88-89
Quark Index Xtention 87, 89-90
QuarkXPress 87
software 87
Encoded Archival Description *see* EAD
encyclopaedias
book indexes 6
entity representation 21
entries 69-70
arrangement 79-80, 191-192
exhaustivity 80-81
form 81-82, 259

layout 186
lead-in 70
main 38-39, 77-78
rating 122-123
recording 78
relationships 78, 123-124
specificity 80-81
sub 39, 70, 77-78
thesauri 173-175
environmental issues 226-232
abbreviations/acronyms 230-231
analytical chemistry 228-229
basic concepts 227
biological sciences 230-231
chemical names 229
environment-specific information 228-230
geographical names 231
Greek letters 230
heavy metals 229
history 226-227
indexing principles 227-228
indexing requirements 228
keywords 228
scientific/popular names 230
social issues 232
species/family names 230
spelling 228
terms 226
vocabulary control 228
WATERLIT 226-227
ephemera
Africana 199-200
equipment needs
web indexing 120
equivalent relationships 177
errors 182
ethics 18
evaluation
antique maps 211-212
indexes 31-37
training 319
events/activities
museum objects 277
events' names
Africana 203
exhaustivity 80-81
expression of concepts 34
eXtensible Markup Language *see* XML

Fairbridge Index 207
fees
web indexing 121
files
web indexing 120, 121
films
Africana 199
final index preparation 82-83
finding aids
archives 216-218
folklore
Khoe-San languages 247-248
format 39-40
formats
periodicals indexes 104-105
free text
museum objects 278
freelance indexers 308
functions
indexes 29-30, 69

geology
book indexes 6
geographical entities
museum objects 276
geographical names 231
global indexes 117
Google 233, 239
government publications
indexes to 5
Grahamstown Journal 4
Greek letters 230
guidelines
abstracting/abstracts 149-154
book indexing 73-74
conceptualisation 22-23
indexing 38-40, 73-74
labour indexing 241
museum objects 270
newspapers indexing 293-295
periodicals indexing 107-110
politics and international relations 302
subject headings 62-63
thesauri 173-175

handwriting
Africana 202
harvesting
metadata 135
headings
form 81-82
standardisation 190-191
health sciences 233-237

book indexing 234
collections 234
Google 233
Index to South African Periodicals 234
indexing aids 234-236
journals indexed 235
major indexes 234-235
Medline 233, 235, 236
MeSH 235-236
National Library of Medicine (Washington) 233
personal files 234
PubMed 233
recent developments 236-237
heavy metals 229
hierarchical relationships 175-176
Historical Papers (University of the Witwatersrand) 218
homographs 81
HTML 122, 124-125, 132
hyperlinking 90-91
Hypertext Markup Language *see* HTML

illustrations
literature for children 261
inconsistencies 41
labour 240
indent style 77, 186
Index to SA Legal Periodicals 252-253
Index to South African Periodicals 4, 92-102, 207, 234
access to 96, 102
additional periodicals 95
application and use 96-97
compilation 93-94, 95-96
contracts 93
database providers 96
focus 94
frequency 101
income generation 96-97
indexing approach 94
information selection 97
ISAPOnline 96, 102
main indexing fields 99-100
National Library of South Africa 96
non-scholarly periodicals 97-99
periodical lists 94-95, 101
periodical selection 94
proposal procedures 95-96
quality control 102
recall 100
research support 97
standards, rules, guidelines 93-94
subject terms 101
support 96-97
technological advancement 102
visibility 102
The Indexer 87
indexers
competencies 14-16, 93, 220, 309-310
freelance 308
intellectual skills 223
literature for children 261-263
in a monetarist world order 18
newspapers 292-293
training 307-322
indexes
Africana 201-202
archives 216-217
author's expectations 38
books 5-6, 68-85
characteristics 29-43
definitions 29-30
electronic 6
evaluation 29-43
format 39-40
formation 34
functions 29-30, 69
government publications 5
health sciences 234-235
length 39
main headings 38-39
newspapers 4-5
periodicals 4
printed 192
publisher's expectations 38
regulator's expectations 38-40
role of 257-258
subheadings 39
subject expectations 37-38
typeface 39
types 31
user expectations 37
indexing 13-14
aids 234
antique maps 210-215
archives 216-220
automatic 16, 35, 111-116
books 68-85
categories 222
in context 7-19
definitions 7-8
difficulties 239-241

education 221-225
environmental issues 226-232
features 292-293
functions 13-14
guidelines 38-40
health sciences 233-237
history 3-6
intellectual skills 223
Internet 17
KWIC 16
KWOC 16
labour 239-242
language 204
law 250-256
levels 105
machine-generated 35
manual 33, 34
metadata 129-140
museum objects 268-284
national bibliographies 289-290
newspapers 292-296
organisation policy 44-49
periodicals 92-110
photographs 218-219
politics and international relations 297-303
postcoordinate 58-60
precoordinate 58-60
principles 227-228
problems/issues 16-18
process 8-9
promotion 241
protocols 218
related disciplines 12-13
requirements 228
resources 201-202
software 83-84
South Africa 3-6
steps 8-9
terms 165, 173, 240, 245-247
training 307-322
verbal subject description 57
web 17, 117-128
indicative abstracts 141-142
information
organisation and retrieval 12, 44, 53-54, 159-160
political economy of 17
retrieval research 10-11
information theory 9-11
Belkin, N 10-11
information retrieval research 10-11
Ingwersen, P 11
informative abstracts 142-143
Ingwersen, P 11
instructional design
training 313-315
International Committee for Documentation *see* CIDOC
International Council of Museums *see* ICOM
international relations *see* politics and international relations
International Standards Organization *see* ISO
Internet 113
indexing 17, 117, 125, 239
intranets 119
introductory notes 72, 190
ISAP *see Index to South African Periodicals*
ISO documents 189, 192-193
item-level metadata 130

Jan Smuts House Library 303
Jewish history
indexes 5
journals *see* periodicals
Juta-National Library Index to the Government Gazette 5

keyword in context indexing *see* KWIC
keyword out of context indexing *see* KWOC
keywords
authors' 107
environmental issues 228
politics and international relations 299-300
keywords meta-tags
web indexing 127
Khoe-San languages 243-249
alphabetisation 246
click sounds 245-247
folklore 247-248
names 245-246
oral data 248
oral history 247-248
ordering of signs 246-247
overview 243-245
recent projects 248
terms 245-247
typestyle 245-246
knowledge systems 17
KWIC indexing 16, 114, 115
KWOC indexing 16

labour 238-242
acronyms 240

alternative press 240
audience 239
guidelines 241
history 238-239
inconsistencies 240-241
indexing difficulties 239-241
indexing promotion 241
Internet indexing 239
mainstream press 241
terms 240
language(s)
Africana 202, 204
automatic indexing 111
indexing systems 34
issues 63-66
Khoe-San 243-249
museum objects 272
verbal subject description 63-66
law 250-256
bibliographic services 252-254
bibliographies 251-252
client services 254-255
context 250-251
Index to SA Legal Periodicals 252-253
South African Legal Abbreviations 253
subject librarians 254
Thesaurus of South African Legal Terms 253-254
Unisa Law Library 250-256
layout
entries 186
run-on 77, 186
set-out 77, 186
learning outcomes
training 316
legal bibliographies 251-252
length
of indexes 39
Lesotho
national bibliographies 285-291
Lesotho: A Comprehensive Bibliography 286-287
Lesotho Annotated Bibliography 288, 290
letter-by-letter arrangement 80, 185-186, 259
levels
metadata 130
of specificity 273
library science 12
linguistics 13
literature for children 257-267
arrangement of entries 259
characteristics of indexes 258-261
cross-references 260
current practice 263-266
format of entries 259
illustrations 261
locators 259-260
media teachers 258
publishers 263-266
role of indexes 257-258
standards for indexes 258-261
subheadings 260
table of contents 261
who should index? 261-263
locators 70, 78-79, 118, 183-184, 191
literature for children 259-260
local indexes 117
London Jewish Chronicle 5

machine-generated indexing 35
macroanalysis 54
main headings 38-39, 69-70
form 81-82
formulating 77-78
mainstream press
labour 241
management
antique maps 210-211
maps *see* antique maps
MARC21 99
matching
automatic indexing 111
materials
museum objects 277
manual indexing 33-34
manuscripts
Africana 200
MDA 270-272
medical indexing *see* health sciences
Medical Subject Headings see MeSH
Medline 233, 235, 236
MeSH 235-236
metacodex records 199
metadata 129-140
administrative 131
collection-level 130
concepts 130
content 131
creating 139-140
crosswalks 135
definition 129
descriptive 131
Dublin Core 135-139
EAD 134

harvesting 135
HTML 132
inter-operability 135
item-level 130
levels 130
purpose 129-130
RDF 133
SGML 132
standards 131-132, 133-134
structural 131
syntax 132
TEI 133
tools 140
types 131
XML 133
meta-tags 118, 125-127
description meta-tag 127
HTML 125-127
keywords meta-tag 127
methods
training 317-319
microanalysis 54, 68
models
training 314, 319-321
modular abstracts 145-147
multi-word headings 182-183
museology 12-13
Museum Documentation Association *see* MDA
museum objects 268-284
cataloguing 268-269
CIDOC Guidelines 270
classification systems 277-278
computer use 269
data standards 270-271
dates 276-277
events/activities 277
fields and structure 271-272
free text 278
geographical entities 276
ICOM 270
language 272
materials 277
MDA 270-272
names 271-275
nomenclatures 273-274
objects vs text 269-270
overview 268-269
people 275-276
specificity 273
styles/periods 277
thesauri 273-274
vocabularies/vocabulary 273-274, 278

national bibliographies 285-291
bibliographic activities 287-289
computer use 288
early bibliography 286
comprehensive bibliography 286-287
the future 290
indexing 289-290
Lesotho 285-291
Lesotho: A Comprehensive Bibliography 286-287
Lesotho Annotated Bibliography 288, 290
place names 290
National Inquiry Services Centre 4, 96, 97, 227, 234
National Library of Medicine (Washington) 233
National Library of South Africa 4, 5, 92, 96
natural language 57, 64-66, 294-295
navigation
web indexing 121-127
needs assessment
training 315
new knowledge systems 17
newspapers 292-296
cross-references 295
guidelines 293-295
indexers 292-293
indexes to 4-5
indexing features 292-293
natural language 294-295
online developments 295
nomenclatures
museum objects 273-274
non-hierarchical relationships 176-177
non-scholarly periodicals
Index to South African Periodicals 97-99
notations 54-57

object names
museum objects 271-272, 275
omissions
abstracts 154
online developments
newspapers 295
online indexing *see* web indexing
oral data collection
Khoe-San languages 248
oral history
Khoe-San languages 247-248
organisation policy 44-49, 80
original text 34

orthography
Africana 203-204

page numbers *see* locators
periodicals
health sciences 235-237
indexes to 4
non-scholarly 97-99
periodicals indexing 92-110
authority files 106-107
authors' keywords 107
classification/coding 107
formats 104-105
guidelines 107-110
Index to South African Periodicals 92-102
indexing levels 105
overview 103-104
principles 103-110
teamwork 110
words from abstracts 106
words from controlled vocabularies 106
words from text 106
words from titles 106
personal files
health sciences 234
personal names
Africana 203
museum objects 275-276
pertinence 37, 113
philosophy 13
photographs
Africana 199
archives 218-219
physical description 54
pitfalls
web indexing 119
place names
Africana 203
national bibliographies 290
political economy of information 17
politics and international relations 297-304
cross-references 301
dates 302
EISA Library 303
elections 298
guidelines 302
indexing rules 299-302
individual approaches 300
international relations 298
Jan Smuts House Library 303
keywords 299-300
neutral stance 300-301
politics 298
thesauri 299
unique features 298-299
postcoordinate indexing 59-60
Boolean operators 59-60
precoordinate indexing 58-59
presentation
book indexes 180-182
printed indexes 192
professionalisation 6
programmes
training 317-321
promotion
labour indexing 241
PubMed 233
publication
antique maps 214
publishers
literature for children 263-266
punctuation 174

qualifiers 184
Quark Index Xtension 87, 89-90
QuarkXPress 87

race relations
book indexes 5
RDF 133
records management 12
reference sources
Africana 202
related disciplines
of indexing 12-13
related terms 34
relationships between entries 78, 123-124
relevance 37
automatic indexing 113
web indexing 122-123
religion
book indexes 5
requirements-oriented approach
conceptualisation 24-25
Resource Description Framework *see* RDF
resources
indexing 201-202
run-on layout 77, 186

SABIB *see South African Bibliography*
SABINET 4, 96, 97, 208
scattering 182

scientific/popular names 230
scope notes
  thesauri 165
search engines 17, 117, 127, 233, 230
services to clients
  Unisa Law Library 254-255
set-out layout 77, 186
SGML 132, 217
short courses/workshops
  ASAIB 310-312
signs
  Khoe-San languages 245-246
simplistic approach
  conceptualisation 24
singular/plural 82, 173
situation analysis
  training 315
social issues
  environmental issues 232
Society of Indexers (United Kingdom) 6, 42
socio-cognitive approach
  conceptualisation 25-26
software 83-84
  embedded indexing 87
  thesauri 177-178
  web indexing 119
South Africa(n)
  bibliography history 205-209
  education 221-222
  indexing history 3-6, 205-209
  literature for children 263-266
  Jewish history 5
South African Bibliography 205-208
  African studies 208
  Africana 201, 202, 206-207
  authority files 208
  developments 206-207
  Fairbridge Index 207
  history 205-206
  Index to South African Periodicals 207
  South African Freedom Struggles 208
  South African National Bibliography 207-208
  South African Studies 208
South African Freedom Struggles 208
South African Legal Abbreviations 253
South African National Bibliography 207-208
South African Studies 97, 208
South Africanisms 204
spaces 187
species/family names 230
specificity 80-81, 273
spelling 174, 228
Standard Generalised Markup Language *see* SGML
standardisation/standards 189-193
  archives 217-218
  arrangement 191-192
  book indexing 73-74
  headings 190-191
  introductory notes 190
  Index to South African Periodicals 93-94
  ISO documents 189, 192-193
  literature for children 258-261
  locators 191
  metadata 131-132
  museum objects 270-271
  presentation 192
structural cataloguing
  antique maps 212-214
structural metadata 131
structure
  thesauri 164-165
structured abstracts 144-145
style
  abstracting 150
  book indexing 76-77
styles/periods
  museum objects 277
subheadings 39, 70
  form 81-82
  formulating 77-78
  literature for children 260
subject
  analysis 17
  description 53-67
  headings 58, 60-63
synonyms 81
syntactic analysis
  automatic indexing 116
syntax
  metadata 132

table of contents
  literature for children 261
task analysis
  training 316
taxonomies
  learning outcomes 316-317
teaching aims/learning outcomes
  training 316-317
teamwork
  periodicals indexing 110

technical presentation
book indexes 179-188
technology 17
TEI 133, 217
telecommunication 13
terminology/terms 165, 166, 173, 223, 224, 226, 240, 245-247
text
searching 112, 113
words from 106, 112
Text Encoding Initiative *see* TEI
thesauri/thesaurus
abbreviations 174
archives 217
authority files 208
characteristics 160-163
compilation steps 169-173
compound terms 174-175
construction of 159-178
definitions 160
equivalent relationships 177
guidelines 173-175
hierarchical relationships 175-176
indexing terms 165, 173
museum objects 273-274
non-hierarchical relationships 176-177
overview 159-160
punctuation 174
relationships application 175-177
relationships between terms 166-169
scope notes 165
singular/plural 173
software 177-178
spelling 174
structure 164-165
use of 163-164, 299
verbal subject description 57-58
words from 106
Thesaurus of South African Legal Terms 253-254
time frames
web indexing 120
titles
words from 106
training 307-320
ASAIB 310-312
teaching aims/learning outcomes 316-317
competencies 309-310
distance education 312-319
evaluation 319
freelance indexers 308
instructional design 313-315
methods 317-319
models 314, 319-321
needs assessment 315
overview 307-308
programmes 317-321
short courses/workshops 310-312
situation analysis 315
skills 312
task analysis 316
taxonomies 316-317
theory 308
web-based instruction 318
what to teach 315
translation 9, 76
trial-run indexing
web indexing 121
truncation 115
typeface 39
types of index 31
typestyle 187

uniform resource locators *see* URLs
Unisa Law Library 250-256
URLs 118
users 37
children 262

verbal subject description 53-67
abstracting 57
controlled vocabulary 66-67
decisions 63
indexing 57
language 63-66
notations versus 54-57
overview 53-54
quality of 67
postcoordinate indexing 59-60
precoordinate indexing 58-59
subject headings 58, 60-63
thesauri 57-58
visual architecture
web indexing 124-125
vocabulary control 64, 66-67
environmental issues 228
museum objects 273-274, 278
thesauri 159

WATERLIT 226-227
web-based instruction
training 318

web indexing 17, 117-128
  authoring tools 119
  availability of files 121
  boundaries 119
  decisions/questions 119-121
  display 119
  equipment needs 120
  fees 121
  HTML 125-127
  input of entries 120
  Internet retrieval 118, 125
  intranets 119
  kinds of files 120
  maintenance 120
  meta-tags 125-127
  navigation system 124-125
  navigational tool 121-124
  overview 117-118
  pitfalls 119
  rating entries 122-123
  relationships between entries 123-124
  relevance 122-123
  software 119
  time frames 120
  trial-run indexing 121
  URLs 118
  visual architecture 124-125
website indexing *see* web indexing
weights/measures/monetary systems
  Africana 204
William Cullen Library (University of the Witwatersrand) 218
Wits Africana Library 214
word association
  automatic indexing 112
word-by-word arrangement 80, 185-186, 259
words
  from abstracts 106
  from controlled vocabularies 106
  from the text 33, 106
  from titles 106
World Wide Web *see* WWW
writing style
  abstracts 150-153
WWW 113, 115, 118, 217

XML 133