

INDEXING FOR SOUTHERN AFRICA

A Manual Compiled in Celebration of ASAIB's First Decade

1994–2004

Editors

Jacqueline A Kalley

Elna Schoeman

Marlene Burger

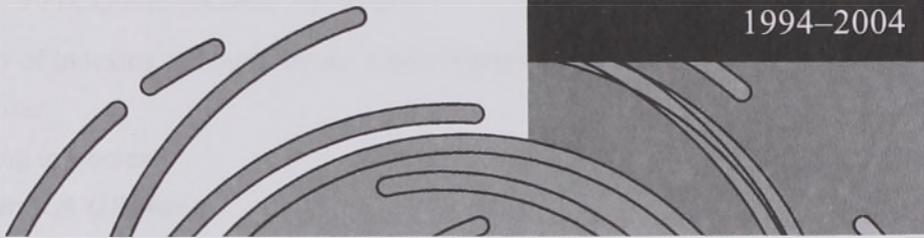


CONTENTS

INDEXING FOR SOUTHERN AFRICA

A Manual Compiled in Celebration of ASAIB's First Decade

1994–2004



Editors
Jacqueline A Kalley
Elna Schoeman
Marlene Burger

UNIVERSITY OF SOUTH AFRICA, PRETORIA

INDEXING FOR SOUTHERN AFRICA

A Journal (Original in English) in Celebration of ASAIB's 25th Birthday



© 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.

UNIVERSITY OF SOUTH AFRICA

CONTENTS

Introduction	ix
<i>JA Kalley</i>	
Part One: THEORETICAL ISSUES	
1 History of Indexing in South Africa: A Brief Overview	3
<i>R Musiker</i>	
2 Indexing in Context	7
<i>JC Theron & GR Davis</i>	
3 Conceptualisation: What is it all About?	20
<i>I Fourie</i>	
4 Characteristics and Evaluation of a Good Index	29
<i>A-M Arnold</i>	
5 Organisation Policy	44
<i>M Burger & I Fourie</i>	
Part Two: GENERAL APPLICATIONS	
6 Verbal Subject Description	53
<i>I Fourie & M Burger</i>	
7 Book Indexing	68
<i>I Fourie & M Burger</i>	
8 Embedded Indexing for Creating Book Indexes	86
<i>E Lastovica</i>	
9 Periodicals Indexing	92
<i>A. Kotze</i>	
Addendum: Principles of Periodical Indexing	103
<i>M Burger</i>	
10 Automatic Indexing	111
<i>P Underwood</i>	

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

Part Three: SPECIAL FIELDS

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

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

Part Four: TRAINING

31	Training Opportunities	307
	<i>M Burger</i>	
	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 1900 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.

1 THE HISTORY OF INDEXING IN SOUTH AFRICA: A BRIEF OVERVIEW



P. MURPHY

ABSTRACT

The history of indexing in South Africa is traced from the early days of the country's independence in 1961 to the present. The article discusses the role of subject terms, the influence of the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biological Sciences (IUBS) and the impact of the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biological Sciences (IUBS) on the development of indexing in South Africa.

Part One

THEORETICAL ISSUES

Introduction

Indexing is a system that was used as the earliest book-orientation approach and, at present, is a logical method of organising books, as in the subscription lists and the use of a hierarchical method for the classification of books. The first historical records of the practice of indexing in the field of chemistry can be traced to the late 18th century, and the practice has since the publication of the *Handbook of Chemistry* by Berzelius and the *Handbook of Chemistry* by Berzelius and the *Handbook of Chemistry* by Berzelius.

Chemistry and indexing were closely related at the time of the early chemical discoveries. The history of chemistry in South Africa has been largely shaped by the development of the country's universities. In 1911, the first university, the University of the Witwatersrand, was established in Johannesburg. It was the first university in South Africa to have a department of chemistry. The first chemistry department in South Africa was established in 1911 at the University of the Witwatersrand in Johannesburg. The first chemistry department in South Africa was established in 1911 at the University of the Witwatersrand in Johannesburg.

As the first university in South Africa, the University of the Witwatersrand played a leading role in the development of chemistry in South Africa. The first chemistry department in South Africa was established in 1911 at the University of the Witwatersrand in Johannesburg. The first chemistry department in South Africa was established in 1911 at the University of the Witwatersrand in Johannesburg.

The University of the Witwatersrand was the first university in South Africa to have a department of chemistry. The first chemistry department in South Africa was established in 1911 at the University of the Witwatersrand in Johannesburg. The first chemistry department in South Africa was established in 1911 at the University of the Witwatersrand in Johannesburg.

The history of indexing in South Africa is traced from the early days of the country's independence in 1961 to the present. The article discusses the role of subject terms, the influence of the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biological Sciences (IUBS) and the impact of the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biological Sciences (IUBS) on the development of indexing in South Africa.

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.



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.



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 (exhaustivity¹)
- the translation of these concepts into words (the issues of vocabulary and specificity²) (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)? <ul style="list-style-type: none"> · 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.

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

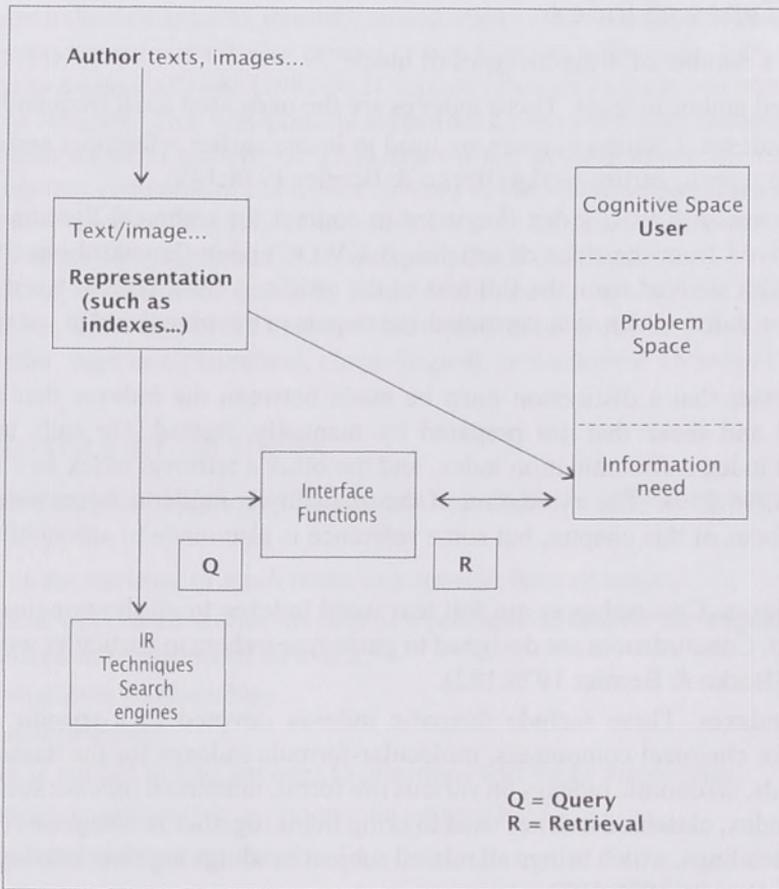


Figure 1: *Indexes as Part of the System and IR Process with the Reader's Needs in Mind* (Ingwersen 1992:55)

- 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.

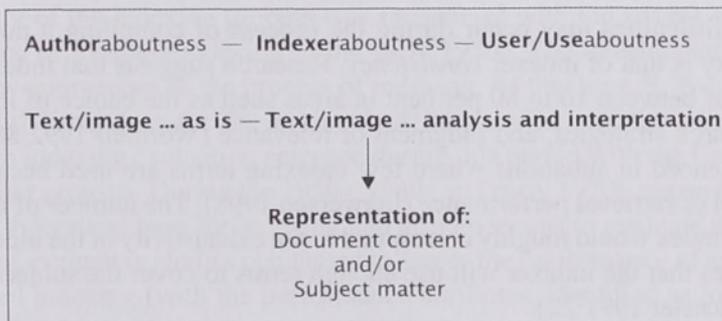


Figure 2: *Elements Involved in the Indexing Process* (Ingwersen 1992:55)

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).

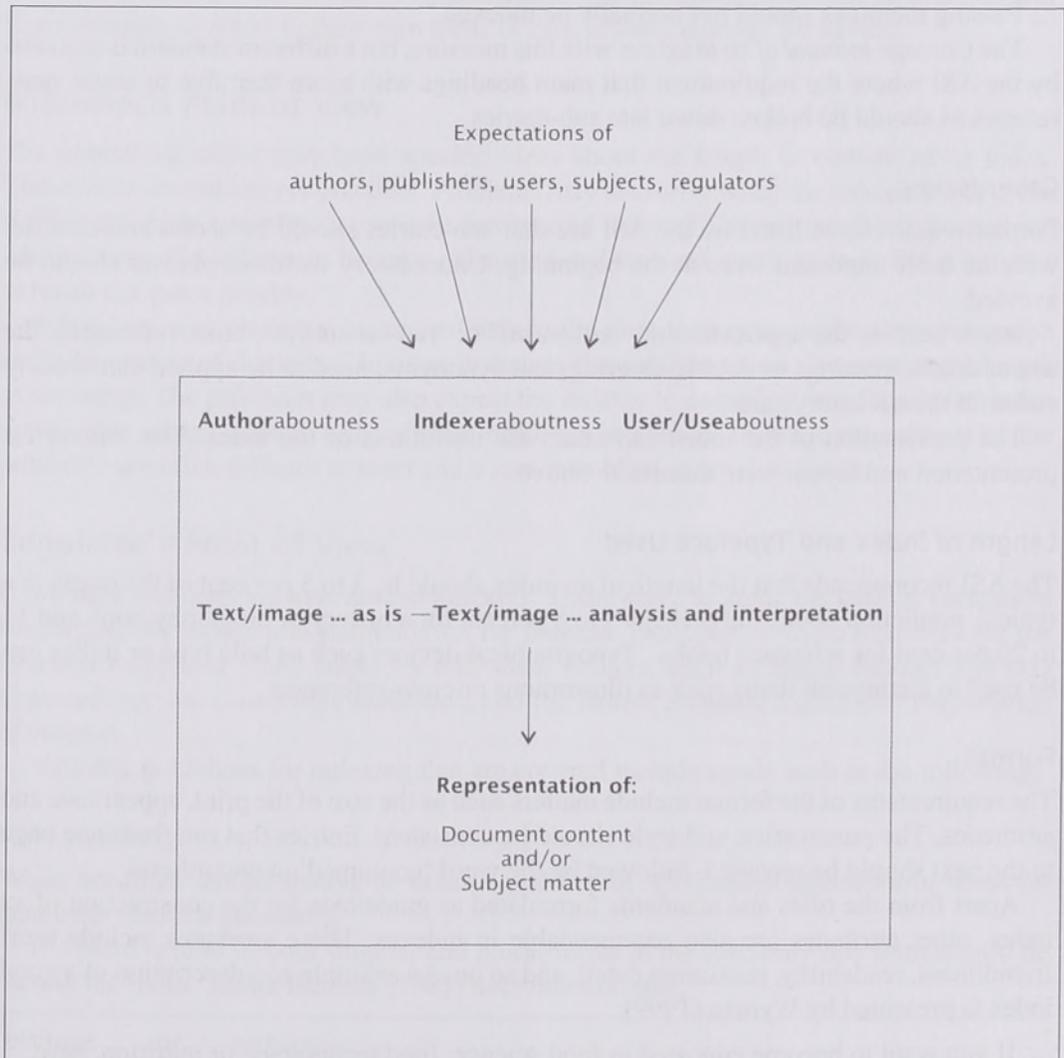


Figure 3: *The Elements Involved in and the Influencing Factors on the Indexing Process* (adapted from Ingwersen 1992:55)

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.

Coverage	What type of work will be included in a good index?
Extent	What is the expected number of entries to be included in the index? How will the index be organized?
Users	Who will use the index? What are their needs?
Available technology	What is the state of the art in indexing technology? What are the strengths and weaknesses of the technology? How can the technology be used to improve the quality of the index?