



Marlene Burger

### *Abstract*

The abstract (produced by machine or human), as a concise statement of the central message of a document, has become an increasingly important tool for distinguishing truly relevant information from the bulk of information available (Pinto & Cálvez 1999:365).

## Abstracting and Abstracts

Abstracting is a verbal method of organising information (i.e. words are used to describe what an entity is about). It is aimed at making the mass of information more manageable so that clients can retrieve relevant information when they need it, and exclude irrelevant information. This means that, when you are writing abstracts for your own information system, you should be guided by your own needs. To be able to do this properly, you should be familiar with the general rules and guidelines that apply to the writing of abstracts.

An abstract is a short, concise and accurate representation (surrogate) of the intellectual content of an entity, its purpose being to promote awareness of the entity and also of information retrieval. Abstracts can be written for entities such as books, reports, journal articles and websites. In this chapter we concentrate on abstracts for journal articles, but the principles can be applied to other types of entity as well.

There are five types of abstract: indicative abstracts, informative abstracts, critical abstracts, structured abstracts and modular abstracts. Diagrammatic abstracts are not discussed. The types of abstract most commonly used are indicative and informative abstracts. Frequently, a combination of indicative and informative abstracts is used.

## Indicative Abstracts

Short abstracts which describe the broad concepts covered in an entity in general terms are called indicative abstracts. General terms like 'are discussed' and 'was investigated' are used to describe the intellectual content of an entity. The purpose of indicative abstracts is merely to show that the information exists: no factual information or findings are included. Indicative abstracts only indicate what is covered in the entity – what is done, what is discussed, what is analysed, what is compared, and so on. Indicative abstracts are presented in a single paragraph only.

Indicative abstracts indicate that important information and/or specific data may be found in a particular entity (e.g. the number of sheep farmers in South Africa is determined according

to types of sheep). Thus indicative abstracts focus attention on the existence of entities. They are considerably shorter than informative abstracts and vary from 50 to 100 words.

Example of an indicative abstract (adapted from Cleveland & Cleveland 1990):

Disraeli, H. 1989. Onion raising in California. *Agriculture Happenings*, 201(2):7-19.

Deals with all aspects of the onion crop in California and touches onion production in other states, mainly for comparative purposes. All types of onions are discussed, with figures given for each category and for the total production. There are breakdowns for how much of the product is consumed in the state and how much is exported. There is also a detailed discussion of farming techniques, wholesale practices, transportation, and general economics of the onion business. A large portion of the article is devoted to the social and medicinal implications of eating onions.

## Informative Abstracts

Longer abstracts that summarise the essential factual content of an entity as concisely and clearly as possible are called informative abstracts. They include *qualitative* and *quantitative* information, as well as the following: the purpose and scope of the entity, the research methods used, the research results, the validity of findings and their applications. This depends on the extent to which the information is covered in the entity – for example not all entities are scientific articles in which these matters are set out step by step. Informative abstracts may be presented in more than one paragraph and vary from 200 to 300 words and may sometimes even be longer. The length of informative abstracts may also be restricted to a tenth or a twentieth of the original entity.

If the information is available, informative abstracts should include the following facts/points about the original entity:

- aim and scope of the entity
- methods of use
- results
- findings
- validity of findings
- applications

Informative abstracts take considerably longer to write than do indicative abstracts. The intellectual content of entities – what they are about – has to be thoroughly analysed first. Knowledge of the subject content is an additional recommendation.

In practice, there is a tendency to write a combination of indicative and informative abstracts (indicative-informative abstracts). This means that the chief aspects are discussed in informative style and the less important aspects in indicative style.

Example of an informative abstract (adapted from Cleveland & Cleveland 1990):

Disraeli, H. 1989. Onion raising in California. *Agriculture Happenings*, 201(2):7-19.

According to this article, California raised 170,000,019 onions while its closest competitors, Oregon and Texas, combined could manage only 70,000,017 onions. Rhode Island was almost out of contention with a little over 2 million. Big, white, sweet onions are the best sellers, with yellow onions second, and small green onions close behind. Some 120 million white onions, 30 million yellow onions, and 20 million little green onions were put on the market. Californians consumed 70,000,035 of these onions, with the rest being placed on the out-of-state market. Most onions farming is highly mechanical and is done by large-scale farming enterprises. The main mode of transportation is by truck, although some is by rail and a minute part by Ford station wagon. The onion business is economically stable and is generally a money-maker. According to the article, persons who eat onions with every meal have a life expectancy of some 1% above the national average, but they are invited to only 20% as many parties, and their divorce expectancy is 200% above the national average.

## Critical Abstracts

Normally abstracts include no criticisms or personal interpretations. Critical abstracts are the exception. Critical abstracts are seldom found in databases with an international target group (e.g. *ERIC* or *Philosopher's index*). Critical abstracts will become increasingly important because of the role they can play in value addition by information services.

Critical abstracts comment, for example, on the

- depth or average of an entity (e.g. a journal article or research report)
- adequacy of the research methodology used and the research findings where applicable
- background needed by the target group (clients) to utilise the entity
- importance of the contributions to the development of knowledge

Although critical abstracts can add value to information services, they are very time-consuming to write and require extensive knowledge of the subject. The abstractor must include a *substantial opinion* about the article. Critical abstracts are similar to critical book reviews – they evaluate the quality of the work and indicate the weaknesses and strengths of the article. This can even include comparisons with other articles.

Example of a critical abstract (adapted from Lancaster 1998):

Pao, YC & Shy, DS. 1982. On the relationship between bulk modulus and relative volume of the lung during inflation-deflation maneuvers. *Journal of Biomechanical Engineering*, 104(2):136-142.

The article presents an equation relating the bulk modulus of the lung to the relative volume during inflation and deflation. The average bulk modulus of the lung was obtained by injecting air via a 6-mm-i.d. cannula in the main lobar bronchus. "Regional lobe" volume changes were measured by roentgen-videographically determined placement of 25 metal markers implanted in the excised lower lobes of three dogs. Whole lobe volumes at various transpulmonary pressures were measured by water

displacement. Pressure and volume measurements were used to calculate bulk modulus ( $K = \Delta V P / \Delta V$ ). The "most satisfactory least squares curve-fit" of bulk modulus ( $K$ ) vs. relative volume ( $V/V_{\max}$ ) was obtained with the equation  $K = C/(1 - V/V_{\max})^n$ . Substituting for bulk modulus, with the equation  $K = VdP/dV$ , and integrating enabled computer-generated pressure-volume plots. This equation provided a better pressure-volume curve-fit than previously obtained, especially at low values pressure and volume. Also, as expected, the bulk modulus was smaller at low volume, but the rate of change of modulus was greater during deflation than inflation.

The authors assumed, without giving sufficient justification, that the "regional lobe" (the area bounded by the 25 markers) included a higher density of airways than the rest of the lobe. Using this assumption, the authors claimed that the modulus and the rate of change of modulus were different for parenchyma tissue and the airways during both inflation and deflation. No mention, however, was made of paired t-tests or any other statistical tests. In fact, if they had done a paired t-test, they would have discovered that none of these differences were significant, even at the 90 percent confidence level.

Other sources of error which were not addressed include: the difference in the properties of excised lung and intact lung due to blood in the vessels, surrounding tissue, negative pressure, etc; the effect of the markers on the pressure-volume relationship; the effect of strain rate on the modulus of lung tissue, which is a viscoelastic material; the time lapsed between regional volume measurement and whole volume measurements (this is important for viscoelastic material); the difference between the true regional  $\Delta V$  and the measured  $\Delta V$ ; and the differences between the mechanical properties of dog and human lung tissue.

Despite its limitations, the article presents a step forward in the understanding of mechanical properties of the lung, and, thus, lung diseases. Therefore, it should be of benefit to researchers interested in respiratory mechanics and physiology.

DS Feldman, USA

## Structured Abstracts

Structured abstracts have the same requirements as informative abstracts, but the content is written in separate paragraphs and under different headings. During analysis of the article, the abstractor must take note of any pre-determined aspects, which will form the headings for the presentation. The following is an example (Lancaster 1998) of a framework for a structured abstract for a specific article:

Irrigation type	Soil type	Crops	Climatic conditions	Place	Results

The abstract need not be presented in columns as illustrated above. The following example (adapted from Lancaster 1998) illustrates an alternative format:

**BACKGROUND:** Structured abstracts, containing several subheadings, have replaced traditional abstracts in most medical journals. Evaluation studies have shown that such abstracts usually provide more information, are of a higher quality, facilitate peer review and are generally welcomed.

**AIM:** The aim of the studies reported here was to investigate an additional possible advantage for structured abstracts, namely whether or not they are easier to search.

**METHOD:** Two studies are reported. In Study 1, using an electronic database, 52 readers were asked to find the answers to two questions for each of eight abstracts in one format (e.g. traditional), followed by two questions for each of eight abstracts in the other format. Time and error date were recorded automatically. In Study 2, using a printed database, 56 readers were asked to find five abstracts that reported a particular kind of study (e.g. studies with school children and reading tests) and then to find five more that reported another kind of study. Again, the order and presentation of the format of the abstracts was counterbalanced. Time and error data were recorded manually.

**RESULTS:** In Study 1, the participants performed significantly faster and made significantly fewer errors using the structured abstracts. There were, however, some unexplainable practice effects. In Study 2, the participants again performed significantly faster and made significantly fewer errors with the structured abstracts. In Study 2, however, there were asymmetrical transfer effects: participants who responded first to the structured abstracts responded more quickly to the following traditional ones than did those participants who responded first to the traditional abstracts.

**CONCLUSIONS:** The overall findings, notwithstanding certain caveats, support the hypothesis that is easier for readers to search structured abstracts than it is to search traditional ones.

## Modular Abstracts

Modular abstracts are intended as full content descriptions of current entities. Each abstract consists of five parts: a citation, an annotation, an indicative abstract, an informative abstract and a critical abstract. This implies that an abstracting service can process it to conform to its own unique requirements with a minimum of effort: any one of the abstracts can be used intact, or edited to form, for example a partially indicative, partially informative abstract, or a partially informative and partially critical abstract. Modular abstracts are mostly accompanied by a modular index and together they represent the concept of 'modular content analysis'.

The prime purpose of modular abstracts is to eliminate the duplication and waste of intellectual effort involved in the independent abstracting of the same entities by several

services. Their standardised format (i.e. the five parts) and treatment also reduce repetitive handling and speed the flow of work within recipient abstracting services.

Example of a modular abstract showing the five parts (adapted from Lancaster 1998):

#### **Citation**

Rosensweig, RE & Beecher, N. 1963. Theory for the ablation of fiberglass-reinforced phenolic resin. *American Institute of Aeronautics and Astronautics Journal*, 1(8):1802-1809.

#### **Annotation**

A theoretical model is developed, for a charring and melting composite material, combining glassy ablation and the char layer-molten glass chemical reaction effects.

#### **Indicative**

The variables associated with the ablation of a typical resin-glass system are examined. These include glass ablation and plastic pyrolysis, flow in both the reacting and non-reacting parts of the melt, mass loss and heat absorption due to chemical reaction. The mathematical development is traced and the approximations utilised are discussed. Parametric examinations are made.

#### **Informative**

Pyrolysis, melting, and chemical reaction are taken into account in this theory of the ablation of phenolic-fiberglass. It postulates a very thin, isothermal, surface reaction zone, where the char layer (carbon) formed during the pyrolysis of the organic binder reacts chemically with the molten silica. Other assumptions are conventional. Calculations for typical IRBM re-entry conditions showed little temperature drop in the reaction zone, 6% maximum and usually less than 1%. Depth of the zone was three orders of magnitude less than the thermal thickness. The unreacting run-off in the melt zone ranged from 40-80% as a function of the possible reaction enthalpy level. However, more than 99% of the material reaching the reaction zone was affected. At the expected temperatures of 1400-2000 °C, the theory assumed the reaction  $\text{SiO}_2 + 3\text{C} > \text{SiC} + 2\text{CO}$ . Earlier experiments had yielded the reaction kinetics. Significant effects, up to 25% increase, on the ablation rate appeared only at the lower reaction rates. Changing the reaction enthalpy by a factor of three changed the ablation rate by less than 10%. When compared with a peak re-entry ablation rate, the value given by this theory was reported to be 38% in defect.

#### **Critical**

This theory extends the classic work of Bethe and Adams (Avco-Everett Research Lab, Res Rept 38, Nov 1958) on ablation of pure glass. Thus it treats the problem as concerning carbon-contaminated glass rather than, as is more usual, a char-layer. In the only comparison given between the theory and experimental data, revealing 38% underprediction by the theory, a thorough error analysis was not included. Spalding (Aero Quart, Aug 1961, pp 237-274) and Scala (General Electric Co (MSVD), Rept R59SD401, July 1959, ARD Jnl, June 1962, pp 917-924) have tested similar problems.

The accompanying modular index entries suggest descriptive terms, drawn from representative index vocabularies that could be used intact, with refinement or with augmentation to index the abstract derived from the modular package. The representative index vocabularies

used as sources for the modular index entries are to be derived from the current indexes or authority lists of the recipient abstracting and indexing services, and will thus reflect the indexing styles and policies of these services.

Example of the modular index entries (adapted from Lancaster 1998) for the modular abstract illustrated above:

### **Physical and Mathematical Systems**

Axisymmetric and Blunt Body Systems  
Re-entry Bodies

### **Environment**

Atmospheric Entry  
Re-entry Conditions  
Space Flight

### **Mass Transfer**

Ablation, Analytical  
Ablation, Charring  
Ablation, Melting  
Ablation of Glasses  
Chemical Reaction Effects  
Gasification Ration  
Reaction Zone  
Thermal Thickness

### **Authors**

Beecher, N  
Rosensweig, RE

### **Thermodynamics**

Carbon-Silica Reactions  
Coupled reactions

### **Materials**

Ablation materials  
Glass Fibers  
Phenolic Resin  
Phenolics, Fiberglass Reinforced  
Reinforced Plastics  
Rocket and Missile Materials  
Thermal (Re-entry) Shields

### **Means and Methods**

Parametric Analysis

### **Affiliations**

Massachusetts Institute of Technology  
National Research Corporation

## Reasons for Writing Abstracts

Authors write abstracts for their own articles, reports, dissertations and theses. You will also encounter abstracts in many of the bibliographic databases, such as *ERIC* and *Library literature*. Abstracts are also used for some full-text databases to enhance the specificity of information retrieval. The reasons why you would write abstracts for your personal information system will be somewhat different from the reasons why abstracts are written as part of publications or for databases intended for general use. For large, general databases, abstracts are written for the following reasons:

- Abstracts make it possible to select entities (e.g. journal articles). You can use the contents of the abstract to decide whether it is worth tracing the complete entity.
- They make clients aware of the existence of specific information (e.g. new research findings, new products or new techniques).
- Abstracts can stimulate new ideas and new interest.
- Well-written and complete informative abstracts can save a great deal of reading time by giving a concise version of the essential facts covered by an entity.
- Abstracts could overcome language problems in instances where entities are available in foreign languages but at least the abstracts are in English. Sometimes an informative abstract may even serve as a substitute if the original entity is not available.
- Abstracts help with the retrieval of information because they offer more search points.
- Abstracts that form part of an entity can help indexers to decide on indexing terms, since these already give a good indication of the contents of an entity.
- Abstracts make it easier to work through the mass of published information and to keep up with new developments.

## Steps in the Writing of Abstracts

Although there are no absolute guidelines or instant recipes for how to set about writing an abstract, there are certain steps you ought to bear in mind when writing indicative and informative abstracts.

1. First study the entity to determine what it is about. This is called *conceptual analysis* and is sometimes a difficult process. There is also no unanimity as to precisely what you should take note of. The complexity of conceptual analysis is covered in detail in chapter 3.
  - Scan the entity at least once to get an idea of what it is about.
  - Note the title, headings and subheadings.
  - Note the introductory and closing paragraphs.
  - Make notes on the key points.
2. On the basis of the conceptual analysis, decide which *key points* covered by the entity should figure in the abstract. Using this as a basis, you then write the abstract according to recognised standards and guidelines. This is referred to as the *translation* of the contents.

- Cover the aim and reason for the research, the scope, the research methods, and the results and findings, if available.
- Compare the points you identified in the conceptual analysis with the reason why the abstract is being written.
- Exclude negative information, generally familiar historical information, points that have not been researched, references to the works of other authors, tables and figures.
- Do not include criticism or your own interpretation.

### 3. Write the abstract, taking the guidelines into account.

- Write the abstract in draft form, referring to the notes you made in step 2.
- The first sentence should convey the nucleus of the intellectual scope of the entity without repeating the title.
- Check whether you have to write an indicative or informative abstract, and whether it would be better to write a combination.
- Note the recommended style for abstracts.
- Note the recommended length.
- Edit the abstract and note the punctuation marks, spelling errors, accuracy, omissions and conciseness. Accuracy is vital.
- Write the final abstract.
- Write an accurate bibliographic description if the abstract does not appear with the original entity. If possible, ask a more experienced abstractor to check your abstract.
- Sign the abstract with your initials.

## Guidelines for Writing Abstracts

There are standards that provide guidelines for the writing of abstracts. For example the following publication:

National Information Standards Organisation. 1994. *Writing abstracts: a draft American national standard*. Bethesda, Md: NISO.

Individual compilers of databases such as *ERIC* create their own guidelines for abstracting according to such standards.

## Bibliographic Description

If the abstract does not appear with the original entity, as in a journal article, it must have an accurate bibliographic description. If the bibliographic description is not accurate, it is very difficult and sometimes even impossible to trace the original entity. There are a variety of guidelines in the subject literature on the format that could be used. Different databases and/or journal publishers may prescribe different formats; but for the purpose of this chapter we follow the Harvard method, since it is an international standard. Keep in mind that most guidelines work with the same data elements even though their order of presentation may differ. These data elements are as follows: title, author(s), publication details, date, page numbers. Examples according to the Harvard method:

DDC21: keeping pace with knowledge. 1996. *Library Association Record*, 96(2):291. Also available online: OCLC Forest Press DDC Web Site <http://www.oclc.or...22ddc/9722book.htm>

Tenopir, C & Jascó, P. 1993. Quality of abstracts. *Online*, 17(3):44-45.

Zin, C. 2000. Success, a structured search strategy: rationale, principles, and implications. *Journal of the American Society for Information Science*, 519(13):1232-1247.

Please note that in some databases the titles of periodicals have standardised abbreviations, that is the full titles are not used. The bibliographic description usually precedes the abstract, appearing at the head of the abstract.

## Abstracting Style

Abstracts should be written in compliance with the style prescribed for the type of abstract, for example indicative or informative abstracts. This implies that one should be able to recognise the type of abstract from the style used.

## Writing Style

Although there are no fixed writing style guidelines, you may heed the following:

- The first sentence of the abstract should give a clear explanation of the intellectual content of an entity, without repeating the title.
- As far as possible, an indicative abstract should always be written in a single coherent paragraph (more paragraphs if informative or critical abstracts).
- It should be written as briefly and concisely as possible – a maximum amount of information should be conveyed with the minimum of words.
- Although you should avoid long, complex sentences, telegram style is not acceptable either. All sentences must be complete. Verbs, articles and prepositions may not be omitted. The first sentence may begin with a verb, for example ‘Describe the state of indexing in South Africa’.
- Use only well-known abbreviations and acronyms. When other abbreviations or acronyms are used for the first time in the abstract, the words must be written out in full followed by the acronym or abbreviation in brackets, for example computer-assisted instruction (CAI). Then, in the rest of the abstract you may use the abbreviation or acronym.
- Any unnecessary words that do not convey significant information should be avoided, such as ‘this article’ and ‘the authors made a study of’.
- As far as possible, the writing style and the sequence of the presentation should correspond to that of the author(s). Be aware that opinions about this do vary and that some authors feel that an abstract should be an independent intellectual creation (this may be achieved when writing a critical or structured abstract).
- Expressions such as ‘I found’ or ‘we found’ should be avoided. Rather use phrases such as ‘findings show’ ‘research indicates’ or ‘the findings are’.

- There are no definite guidelines on the use of the past or present tense. You should be guided by what sounds best.

The following is a list of possible words and phrases that can be used when writing abstracts:

Acknowledges  
 Addressed by  
 Addresses the problem/issue of  
 Advances arguments for/against  
 Advises  
 Advocates  
 Analyses  
 Argues that  
 Asserts that  
 Assesses

Brief discussion of  
 Briefly mentions/looks at

Challenges  
 Classifies  
 Commands  
 Comments on  
 Compares the ... with ...  
 Concentrates on  
 Concludes that  
 Critically evaluates  
 Criticises

Debates the  
 Defines the concepts  
 Delineates  
 Describes  
 Differentiates between  
 Discusses  
 Distinguishes

Elucidates  
 Emphasises  
 Examines  
 Explains  
 Explanation of  
 Explores  
 Expresses concern for

Focuses on  
Forecasts future developments in  
Foresees

Highlights

Identifies  
Illustrates  
In-depth discussion of  
Includes  
Introduces  
Investigates

Launches a strong attack on  
Lists the winners of  
Literary criticism of

Maintains that

Obituary of  
Offers guidelines for  
Outlines  
Overview of

Pays tribute to  
Pleads for  
Popular discussion  
Profile of

Questions whether/the  
Quotes

Reasons for  
Recommends that  
Reflects on  
Refutes  
Regrets  
Rejects  
Remarks on  
Reprint of extracts from  
Review

- Review of a book which
- The reviewer contends/implies/maintains that
- The reviewer describes the book as

- Applicable only to review articles

Scrutinises  
Shows how  
Studies  
Study of  
Suggests  
Summarises  
Summary of  
Surveys

Tabulates data on  
Traces the career/history of

Wide-ranging discussion on

Bellardo (1991:34) gives the following examples to promote the economical use of words:

Phrase	Suggested alternative
because of the fact that	because
consensus of opinion	consensus
for the purpose of	for
in the event that	if
at this point in time	now
at that point in time/during the period of time that	then/while
in order to	to
in a similar manner	similarly
in view of the fact that	since/because

## Accuracy

Accuracy is a very important requirement, especially in informative abstracts that may sometimes be used to replace the original entity. The accuracy of numerical data in particular and of names should be verified. For instance, there is a big difference between 1996 and 1696. If the author states that something ought to be researched or is going to be researched, you cannot say that it has already been researched. Nor should the author's emphasis of certain points be changed.

## Comprehensiveness

An abstract must encompass all the main points covered in the entity. If the abstract is to be used for information retrieval, the aim and coverage of the information system, such as the database, should be taken into account. No concept or point contained in the entity should, however, be discussed in greater detail than was done by the author in the original entity. This also applies to the analysis of tables and graphs.

## Omission of Information

The following information should preferably *not* be included in an abstract:

- References to other entities or contributions by other authors. Such references are permissible only if the original entity consists mainly of a discussion of another entity, or when you are writing a critical abstract.
- References to specific tables, graphs and figures. Remember that the table, graph or figure is not included in the abstract. This rule does not apply for critical or structured abstracts.
- Negative information such as ‘the needs of information practitioners were not researched’: research that has not been done is not mentioned.
- Criticism or interpretation. This may only be included in critical abstracts or other types of abstract with a special function.
- Historical or background information.
- Recommendations about future research.
- References to quotations in the original entity.
- Unfounded comments, well-known facts, speculation or information that does not appear in the original entity.

## Signing the Abstract

Normally abstracts are initialled by the abstractor. In critical abstracts the abstractor’s full name and credentials may be required.

## Automatic Abstracting

Since the purpose of this chapter was to write about abstracts in general only, this section is but a brief discussion of automatic abstracting. In automatic abstracting computers are used to prepare extracts or summaries of the electronic text of entities. Lancaster (1998:267) states the following about automatic abstracting: ‘If computers can be programmed to select terms from documents according to frequency criteria [as in automatic indexing], they can also be programmed to select sentences from documents.’ This is the basis of what is usually called ‘automatic abstracting’ although it is more accurately referred to as ‘automatic extraction’.

Automatic abstracting involves deriving meaningful sentences from the text of an entity, and compiling these sentences to reflect the intellectual content of the entity. Apart from word frequencies, the computer program also notes the relationships between the keywords in a sentence. (If you are interested in finding out more, please read the article by Luhn.) The style of an entity may also be used as a starting point in compiling abstracts. Here the first and last sentences of each paragraph are considered, or the first sentence of each paragraph only. Subject-oriented thesauri have also been used to identify important words in sentences. However, no real success has yet been achieved with automatic abstracting – it cannot replace human abstracting. Sometimes it supplies concise, understandable information, but too often the abstracts read like strings of disconnected sentences, which indeed they are. The main problem with abstracting by sentence extraction is that there is a lack of balance and cohesion. In experiments the following principles have been used:

- Sentences have a greater weighting if certain keywords appear.
- Certain words serve as cue words. If these words appear in a sentence, it implies that the sentence deals with a certain aspect.
- Words appearing in the title and headings of sections or subheadings, give a good indication of what an entity is about.
- Weighted words are assigned in accordance with the position of sentences. Sentences found in the introductory or concluding paragraphs will have a greater weighting than those appearing elsewhere in the entity.
- Sentences are rejected if certain words appear in them.
- Identifying sentences that are good indicators of the content, for example sentences that contain 'the main idea is' or 'a method is described'.
- Parsing of texts – the grammatical parts of a sentence are examined and identified.

The more formal the text, the easier it is to apply automatic abstracting. Johnson (1995:32) explains automatic abstracting as follows: 'The lack of coherency in automatic abstracts, where the reader is left to infer the relationship between the ideas expressed, detracts from the meaning conveyed and requires considerable time and effort for its understanding'. Much more research is therefore necessary to refine the process of automatic abstracting. A comparison between manual and automatic abstracting reflects the following (adapted from Lancaster & Warner 1993:274):

**Manual abstracting:** Select key sentences; paraphrase key concepts/sentences; organise into coherent paragraph; result: abstract.

**Automatic abstracting:** Select sentences by highest numeric score; modify sentences by linguistic analysis; organise sentences by rank statistical order or by sequence in which they occur; organise sentences by structural analysis; result: extract.

While Luhn (1958) and other authors used word phrase frequency in the selection of sentences, other investigators have proposed or used alternative criteria. Edmundson

(1969) identified four possible methods: the key method, the cue method, the title method and the location method. (For more information on these methods, please read the article by Edmundson.)

An example of a Luhn auto-abstract as it appears in Lancaster (1998:268) illustrates the key method:

*Source: The Scientific American, Vol. 196, No. 2, 86-94, February, 1957.*

*Title: Messengers of the nervous system*

*Author: Amedeo S. Marrazi*

*Editor's Sub-heading: The internal communication of the body is mediated by chemicals as well as by nerve impulses. Study of their interaction has developed important leads to the understanding and therapy of mental illness.*

**Auto-abstract\***

It seems reasonable to credit the single-celled organisms with a system of chemical communication by diffusion of stimulating substances through the cell, and these correspond to the chemical messengers (e.g., hormones) that *carry* stimuli from cell to cell in the more complex organisms (7.0)†

Finally, in the vertebrate animals there are special glands (e.g., the adrenals) for producing chemical messengers, and the nervous and chemical communication systems are intertwined: for instance, release of adrenalin by the adrenal gland is subject to control both by nerve impulses and by chemicals brought to the gland by the blood (6.4)

The experiments clearly demonstrated the acetylcholine (and related substances) and adrenalin (its relatives) exert opposing actions which maintain a balanced regulation of the transmission of nerve impulses (6.3)

It is reasonable to suppose that the tranquilizing drugs counteract the inhibitory effect of excessive adrenalin or serotonin or some related inhibitor in the human nervous system (7.3)

\* Sentences were selected by statistical analysis when they had a degree of significance of 6 and over.

† Significance factor is given at the end of each sentence.

Kuhlen (1984:98) states that we do not know enough about the intellectual processes involved in abstracting to develop programs whereby this activity can be simulated by computer: 'Abstracting ... is an intellectual art and as such not directly transferable to automatic procedures. Cognitive psychology and artificial intelligence have, so far, not provided us with sufficient knowledge about the processes really going on in abstracters' minds when they understand texts and condense them. Thus the direct imitation of an intellectual procedure such as abstracting seems to be out of reach.'

A substantial amount of research and development in this regard has been undertaken during the past few years, but Kuhlen's words still seem relevant today.

## Conclusion

While some progress has been made in applying computers to various tasks related to information retrieval, there is little evidence that automatic procedures can as yet outperform humans in such intellectual tasks as indexing, abstracting, thesaurus construction and the creation of search strategies. However, computers can be used to improve the productivity of abstractors and the quality of their work. It is most definitely an area where new developments should be monitored.

## BIBLIOGRAPHY

- Bellardo, T. 1991. *Subject indexing: an introductory guide*. Washington, DC: Special Libraries Association.
- Cleveland, DB & Cleveland, AD. 1990. *Introduction to indexing and abstracting*. Englewood, Colo: Libraries Unlimited.
- Craven, TC. 1993. A computer-aided abstracting tool kit. *Canadian Journal of Information and Library Science*, 18(2):19-31.
- Cremmins, ET. 1996. *The art of abstracting*. 2nd ed. Arlington, Va: Information Resources.
- Edmundson, HP. 1969. New methods in automatic extracting. *Journal of the Association for Computing Machinery*, 16(2):264-289.
- Hartley, J. 1999. From structured abstracts to structured articles: a modest proposal. *Journal of Technical Writing and Communication*, 29(3):255-270.
- How to implement a naturalistic model of abstracting: four core working steps of an expert abstractor. 1995. By B Endres-Niggermeyer ... et al. *Information Processing and Management*, 31(5):631-647.
- Johnson, FC. 1995. Automatic abstracting research. *Library Review*, 44(8):28-36.
- Kuhlen, R. 1984. Some similarities and differences between intellectual and machine text understanding for the purpose of abstracting, in *Representation and exchange of knowledge as a basis of information processes*, edited by HJ Dietschmann. Amsterdam: North-Holland:87-109.
- Lancaster, FW. 1998. *Indexing and abstracting in theory and practice*. 2nd ed. London: Library Association.
- Lancaster, FW & Warner, AJ. 1993. *Information retrieval today*. Arlington, Va: Information Resources Press.
- Liu, J, Wu, Y & Zhou, L. 1999. A hybrid method for abstracting newspaper articles. *Journal of the American Society for Information Science*, 50(13):1234-1245.
- Luhn, HP. 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2):159-165.
- O'Connor, BC. 1996. *Explorations in indexing and abstracting: pointing, virtue, and power*. Englewood, Colo: Libraries Unlimited.
- Pinto, M & Cálvez, C. 1999. Paradigms for abstracting systems. *Journal of Information Science*, 25(5):365-380.
- Pinto, M & Lancaster, FW. 1999. Abstracts and abstracting in knowledge discovery. *Library Trends*, 48(1):234-248.
- Tenopir, C & Jascó, P. 1993. Quality of abstracts. *Online*, 17(3):44-55.

- Wellish, HH. 2000. *Glossary of terminology in abstracting, classification, indexing and thesaurus construction*. Medford, NJ: Information Today.
- Wheatley, A & Armstrong, CJ. 1997. Metadata, recall, and abstracts: can abstracts ever be reliable indicators of document value? *Aslib Proceedings*, 49(8):206-213.



Marlene Burger

### *Abstract*

A thesaurus is a formally structured aid that is used to index and retrieve entities. Although thesauri are very useful, they also create problems that cannot always be solved. With the development of the Internet there has been renewed interest in thesauri. It is therefore necessary to understand the basic principles of thesaurus construction.

### Introduction

Subject approaches in the electronic era have become a major way of finding information. Search engines are trying to fill the void on the Internet, yet users are becoming increasingly frustrated with the thousands of 'hits' from keyword searches (Taylor 2004:261):

With the massive increase in availability of recorded information it becomes more and more evident that keyword searching alone will not suffice. Virtually every word in the English language has more than one meaning or sense, and many of those senses have more than one nuance; many words can be used as nouns, verbs, adjectives, or adverbs. Search systems that purport to allow the user to use "natural language" cannot yet successfully distinguish among meanings or parts of speech in very large general systems, although progress has been made in narrow subject areas.

We also know that authors writing about the same concepts often do not use the same words to express them, and users searching for the same concept do not think of the same words to search for it. The implication is that controlled vocabulary is needed to reconcile all the various possible words that can be used to express a concept and to differentiate among all the possible meanings that can be attached to certain words.

The purpose of this chapter is to introduce you to another aid to vocabulary control, namely thesauri. Thesauri are used in practice to retrieve information (e.g. from databases) or to index entities (e.g. periodical articles). You may even be required to compile a thesaurus for an internal database or for sources on the Intranet. Thesauri are also used in museums and archives. Information professionals may have doubts about the future role and value of thesauri; when compared to natural language or information organisation and retrieval systems (IOAR systems), there does not seem to be any need for thesauri, but with the Internet and information being freely available, more attention is being given to the compilation of thesauri for use on the Internet and to thesauri that can be used to retrieve information.

Aitchison, Gilchrist and Bawden (1997:xiii) explain it as follows: 'Other authors (with whom we agree) maintain that far from being dead, the thesaurus will continue to be a useful adjunct to retrieval systems, though increasingly hidden from the user ... Certainly it is the case that around the world there are many systems that continue to use thesauri for controlled indexing, albeit many such systems offer free text searching as well.' Aids for vocabulary control solve some problems associated with natural language, but they also create problems of their own.

You will come across a variety of definitions of a thesaurus in the literature. It is generally accepted that a thesaurus is formally structured to show the relationships between terms, and it serves as an aid to indexers and others who use IOAR systems to retrieve information. A thesaurus is therefore an aid to be used when *organising* and *retrieving* information. Thesauri can be arranged alphabetically or hierarchically, they can be used for a computerised database or a printed index, and they are available in printed or electronic formats (online thesauri). The National Information Standards Organization (1994:1) explains a thesaurus as follows: 'A thesaurus, for the purpose of this standard, is a controlled vocabulary of terms in natural language that are designed for postcoordination. The need to control the information and use of terms stems naturally from the two basic features of natural language, namely synonyms (different terms representing the same concept) and polysemes or homographs (terms with the same spelling representing different concepts).'

Walker and Janes (1998:305) provide the following definition: 'An alphabetical listing of subject words that comprise the controlled vocabulary for a particular database. Terms from the thesaurus are selected by indexers to describe the document when it is being input, and by searchers to find documents for retrieval. Most thesauri also list subject-related terms for each entry as an aid to selecting the most specific terms.'

There is a different thesaurus for virtually every subject one can think of. The following figure (Figure 1) is an extended example (adapted from Lancaster & Warner 1993:8) of an IOAR system and the role of vocabulary. Note that there are a number of aids for vocabulary control, such as classification systems, thesauri and subject heading lists.

## Characteristics of Thesauri

The following are characteristics of thesauri (more information on these characteristics later in the chapter):

- Are formally structured (e.g. alphabetically or hierarchically).
- Indicate a variety of indexing terms that the indexer/user might think of. Some of these terms are used as preferred terms in the thesaurus, while others are non-preferred terms or entry terms.
- Refer indexers/users from the terms not suitable for indexing (i.e. the entry terms) to the preferred terms used in the thesaurus.
- Indicate relationships with other terms. These relationships are indicated by means of special codes, such as BT (broader terms), NT (narrower terms), RT (related terms).

- Sometimes explain the meaning of indexing terms or how the thesaurus uses the term. This is generally indicated with the code SN (scope notes).

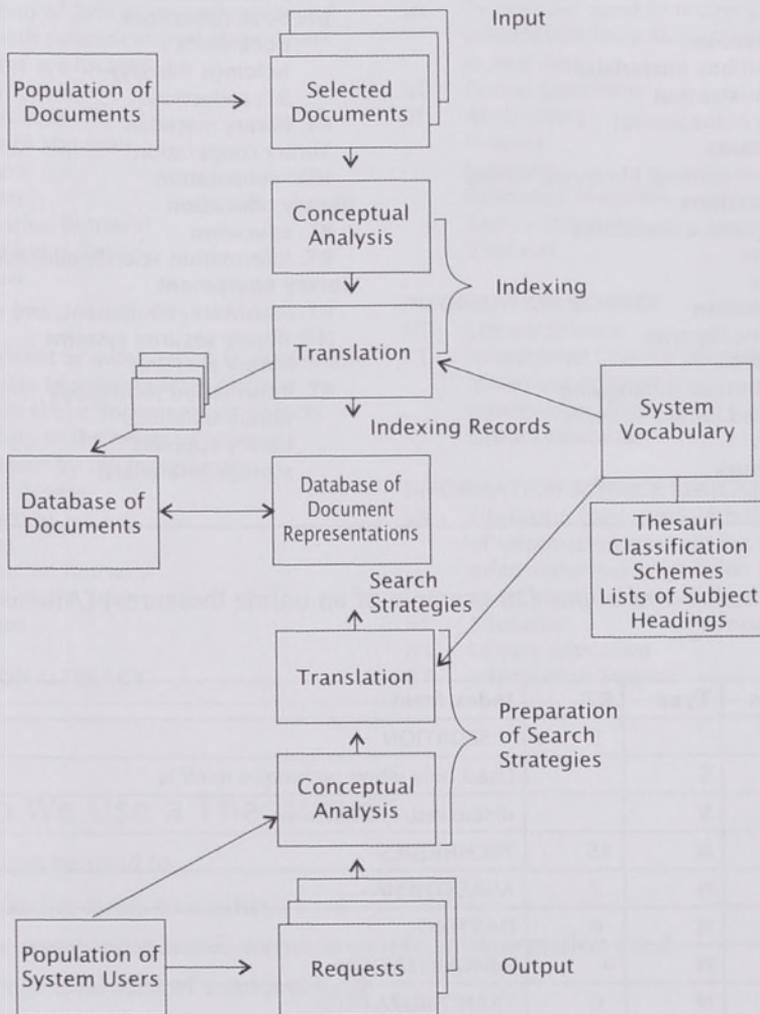


Figure 1: Example of an IOAR system

The following are examples (extracts) from different thesauri.

From the *Asis thesaurus of information science and librarianship* (1994:48):

## ALPHABETICAL DISPLAY

### library access

BT: access to resources  
RT: information access  
libraries

### library acquisitions

USE: acquisitions (materials)

### library administration

USE: library management

### library assistants

USE: paraprofessional library personnel

### library associations

BT: professional associations  
RT: librarians  
libraries

### library automation

UF: electronic libraries  
BT: automation  
NT: computerized cataloguing  
RT: integrated library systems  
libraries

### library buildings

BT: buildings

RT: libraries

library equipment

### library collections

UF: book collections  
book stock  
holdings (library)  
BT: collections

RT: library materials

library cooperation

USE: cooperation

### library education

BT: education

RT: information science education

### library equipment

BT: (hardware, equipment, and systems)  
NT: library security systems  
library shelving  
RT: information technology  
library buildings  
library supplies  
storage (materials)

From *Zoological record online* (an example of an online thesaurus) (Aitchison, Gilchrist & Bawden 1997:92):

Ref	Items	Type	RT	Index form
R1	24		9	* SEDATION
R2	0	S		Used only when technique itself is
R3	0	S		discussed/evaluated
R4	46	B	45	TECHNIQUES-
R5	0	N	7	ANAESTHESIA-
R6	0	N	6	DARTING-
R7	0	N	4	IMMOBILIZATION-
R8	0	N	6	TRANQUILIZATION-
R9	4	R	10	HANDLING-TECHNIQUES
R10	1	R	6	PHYSICAL RESTRAINING TECHNIQUES

In the thesaurus entry, the term entered is listed first. If scope notes are included, they are listed next, indicated by S under Type. Broader terms are listed under Type as B (for examples, TECHNIQUES-), narrower terms are listed under Type as N (DARTING-), and related terms as R (HANDLING-TECHNIQUES). At this point, users can select the appropriate term and search or explode the term. To explode, use the term, including hyphens, followed by an exclamation mark (!). This will search the term and all narrower terms listed under it.

From the *Thesaurus of ERIC descriptors* (1995) (some entries shortened):

<b>Index terms</b>	RT Library Instruction
USE Subject Index Terms	Search Strategies
	Users (Information)
<b>INDEXES</b>	<b>INFORMATION RETRIEVAL</b>
SN Systematic guides to information, consisting of lists of logically arranged items with references that show where the items are located	SN Techniques used to recover specific information from large quantities of stored data
UF Concordances	NT Online Searching
NT Citation Indexing	RT Abstracting
BT Reference Materials	Indexes
RT Abstracts	Indexing
Indexing	Reference Materials
Information Retrieval	Search Strategies
Subject Index Terms	Thesauri
Thesauri	
<b>INDEXING</b>	<b>INFORMATION SCIENCE</b>
SN Assignment of index terms to documents or objects in order to later retrieve or locate these documents or objects according to the selected concepts designated by the index terms	NT Library Science
UF Subject Access	RT Information Science Education
RT Abstracting	Information Scientists
Indexes	Information Skills
Information Retrieval	Library Education
Subject Index Terms	
Thesauri	<b>INFORMATION SCIENCE EDUCATION</b>
	SN Education concerned with the handling of information, comprising such information-oriented fields as Computer and Communication Science ...
<b>INFORMATION LITERACY</b>	BT Education
	NT Library Education
	RT Information Science

## When Do We Use a Thesaurus?

A thesaurus can be used to

- assign indexing terms to entities
- determine appropriate search terms to satisfy an information need
- promote a general idea of a subject field
- offer a semantic map of the interrelationships between terms
- help provide definitions of terms

Generally we only use a thesaurus when searching for information; we call this type of thesaurus a search thesaurus. A search thesaurus can help a user to identify search terms and to think of more general or specific terms, synonyms and related terms. The purpose of a thesaurus is therefore to

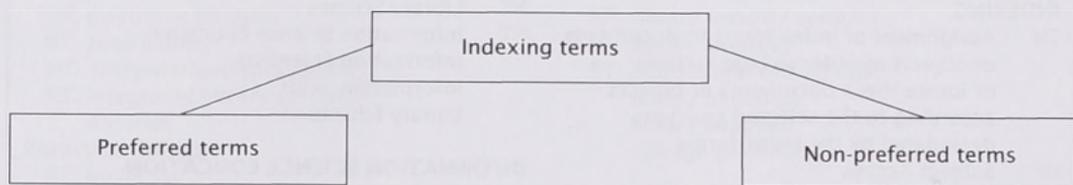
- indicate what terms may be used in an IOAR system

- indicate what terms may not be used in an IOAR system
- suggest terms to the indexer or person seeking information under which further cross-references can be made or information sought and thus promote consistency in indexing
- ensure that a topic is always described in the same way so that information is not lost because different words or phrases are used to describe the same topic
- indicate hierarchical and other relationships between terms

## Structure of a Thesaurus

Thesaurus entries consist of indexing terms that include

- preferred terms
- forbidden or non-preferred terms, thus



A typical entry for a preferred term is as follows:

INDEXING TERM	
SN	(scope note)
UF	(used for)
BT	(broader term)
NT	(narrower term)
RT	(related term)
LIBRARIES	
SN	Institutions housing collections of systematically acquired and organised information resources, and usually providing assistance to users
UF	Resource centres
BT	Institutions
NT	Academic libraries
RT	Library funding

Not every indexing term refers to a broader, narrower or related term; it depends on the nature of the indexing term. Under each indexing term the indexer/user will be referred to other terms that are important. Some terms do not indicate relationships with other terms and these are called 'orphan terms'. A typical entry for a non-preferred term is as follows:

Entry term  
 U (use reference)  
 Resource centres  
 U **LIBRARIES**

The entry term directs the indexers/users from the terms not suitable for indexing to the terms preferred by the thesaurus (and also the IOAR system). Please note the use of font and case to differentiate between terms – that is bold, and capital and small letters.

## Indexing Terms

Indexing terms are also referred to in the thesaurus as keywords, catchwords, descriptors and identifiers. An indexing term, which can comprise one or more words (try to avoid phrases), represents an idea or concept. Guidelines for formulating indexing terms are in a further section of this chapter.

## Scope Notes

Indexing terms do not need to have a scope note. They are only used to explain or describe the meaning of a term if necessary. A scope note can be used

- to indicate the limitations of the meaning of a term

### **INCOME**

SN Income of individual organisation or person. Otherwise use 'National Income'

- to indicate a series of subjects to be covered

### **MECHANICAL COMPONENTS**

SN Includes manipulating parts, separating parts, linkages, internal parts, seals and gaskets, bearings, fasteners, chains, etc.

- to give instructions to indexers

### **INTERIOR LIGHTING**

SN For lighting buildings or spaces, combine with appropriate terms, e.g. 'Shop lighting' use 'Shops' and 'Interior lighting'

Scope notes are sometimes used to explain terms that have the same spelling but different meanings. For these we use qualifiers, for example

cranes (birds)  
 cranes (lifting equipment)  
 mercury (metal)  
 Mercury (planet)  
 Mercury (Roman deity)

The scope note is not part of the indexing term. *A qualifier, however, is part of the indexing term.*

## Relationships between Terms

We distinguish between three types of relationships between terms:

- equivalent relationships (USE and UF references are mostly used)
- hierarchical relationships (BT and NT are usually used to indicate broader and narrower terms respectively)
- non-hierarchical relationships (RT is generally used to indicate related terms)

<b>Equivalent relationships</b> U, UF for 'use' and 'used for' are used to indicate equivalent relationships	
<b>True synonyms</b>	<b>Examples</b>
Trade names	<b>COCA COLA</b> UF Coke Coke U <b>COCA COLA</b>
General use vs scientific use	Rabies or Hydrophobia
Local term vs international term	Aerial (UK) or Antenna (USA)
Competitive term for a new concept	Teleworking or Distance working
Standard terms vs slang	High Fidelity equipment or Hi-Fi equipment
Older vs newer terms	Wireless or Radio
Different stems	Terrestrial magnetism or Geomagnetism
<b>Quasi-synonyms</b>	<b>Examples</b>
Terms representing different poles of the same characteristic	<b>STABILITY</b> UF Instability Instability U <b>STABILITY</b>
Terms that overlap a great deal	Genetics / Heredity Accuracy / Precision
Specific concepts falling under one broad term, if the individual term is regarded as too specific	<b>CEREAL CROPS</b> UF Barley Rye Wheat
Various spellings of a term	<b>COLOUR</b> UF Color Color U <b>COLOUR</b>
Abbreviations	<b>VISUAL DISPLAY UNITS</b> UF VDU VDU U <b>VISUAL DISPLAY UNITS</b>

The person responsible for compiling the thesaurus will have to decide which terms will be the preferred terms and which the non-preferred terms. The thesaurus to be used by the indexer/user might look like this:

**AERIAL**

UF Antenna (when British terms are preferred)

Antenna

U **AERIAL****COCA COLA**

UF Coke (a choice must be made between the two terms)

Coke

U **COCA COLA**

Geomagnetism

U **TERRESTRIAL MAGNETISM**

Hydrophobia

U **RABIES****RABIES**

UF Hydrophobia (a choice must be made between the two terms)

**RADIO**

UF Wireless (because this is the older of the two terms)

**TERRESTRIAL MAGNETISM**

UF Geomagnetism (a choice must be made between the two terms)

Wireless

U **RADIO** (because this is the newer of the two terms)

The list of examples listed under quasi-synonyms can be presented in a similar way. Further relationships between terms can be indicated when it is necessary. For example:

**RABIES**

UF Hydrophobia

BT Animal diseases  
Diseases

RT Treatment

No further relationships must be indicated for entry terms:

Hydrophobia

U **RABIES**

We will now look at another type of relationship, namely hierarchical relationships. There are many more examples and fine divisions, but these examples should suffice for this introductory chapter on thesaurus construction. For more information, please consult the bibliography at the end of this chapter.

<b>Hierarchical relationships</b> BT (broader or more general term) and NT (narrower or more specific term) are generally used to indicate hierarchical relationships	
<b>Genus/species relationships</b>	<b>Examples</b>
<i>This relationship is a link between the class or category and the members of the species. When one has to decide on narrower terms, always ask: Is it a type ...? If not, then it is a related term. For broader terms, always ask: Which is ... a type?</i>	Dogs (genus)                      Poodles (species) Viral diseases                      Measles (species) Libraries (genus)                      Special libraries (species)  <b>DOGS</b> <b>POODLES</b> NT Poodles                                      BT Dogs
<b>Part/whole relationships</b>	<b>Examples</b>
<i>System and organs of the body</i>	Ear (whole)                                      Middle ear (part)  <b>EAR</b> <b>MIDDLE EAR</b> NT Middle ear                                      BT Ear
<i>Geographical location</i>	Africa (whole) Southern Africa (part) Economy (whole) Commercial economy (part)
<i>Hierarchical social structure</i>	Methodist Church organisation (whole) Methodist district (part) Methodists (circuit) (part)
<b>Instance relationships</b>	<b>Examples</b>
<i>This is the relationship between a general category of things and happenings, and individual examples</i>	Sea (category) Baltic Sea (example) Black Sea (example) French cathedrals (category) Chartres Cathedral (example) Rouen Cathedral (example)
<b>Poly-hierarchical relationships</b>	<b>Examples</b>
<i>Some terms occur in more than one hierarchy</i>	Nurses (genus) Nursing administrators (species) Health administrators (species) Nursing administrators (species)  <b>NURSING ADMINISTRATORS</b> BT Health administrators Nurses  <b>ZAMBIA</b> BT Africa Southern Africa

For each of these examples, entries are made in the thesaurus and other relationships are indicated where necessary. We will now look at non-hierarchical relationships. There are many more examples and fine divisions, but these examples should suffice for this introductory chapter on thesaurus construction. For more information, please consult the bibliography at the end of this chapter.

<b>Non-hierarchical relationships</b> RT (related term) is generally used to indicate non-hierarchical relationships. The purpose of related terms is to remind the user who is looking at one term, that he or she may also be interested in another order to retrieve relevant information	
	<b>Examples</b>
Species of the same genus	<b>SPECIAL LIBRARIES</b> RT University libraries  Both are species of the genus LIBRARIES
Object/Part	<b>BOOKS</b> RT Book covers
Object/Attribute	<b>INFORMATION SYSTEMS</b> RT Reaction time
Object/Process	<b>JOURNALS</b> RT Journal catalogues
Object/Application	<b>COMPUTER DATABASES</b> RT Information retrieval
Attribute/Process	<b>EFFECTIVITY</b> RT Effectivity measurement
Process/Object	<b>CATALOGUING</b> RT Catalogues
Process/Attribute	<b>INFORMATION RETRIEVAL</b> RT Precision

For each of these examples entries are made in the thesaurus and other relationships are indicated where necessary.

## Compiling a Thesaurus

When constructing a thesaurus, there are a number of steps you should work through. These steps as well as the standards applying to thesaurus construction are discussed in the following paragraphs.

### Step 1: Analyse why a Thesaurus is Necessary

You might need a thesaurus to index entities for an international database such as *ERIC*, *International information system for nuclear science*, or a local database about South African literature in library and information science, or to retrieve information from a database on the WWW. To do this, you need a sound knowledge of the IOAR system where the thesaurus will be used. It is also important to weigh up the merits of controlled vocabulary and natural language in the situation where the IOAR system will be used.

## Step 2: Check if there Are Existing Thesauri that Can be Used or Adapted

Lists of thesauri are readily available. Aslib publishes an updated list every six months in the *Current Awareness Bulletin*. Some issues of *Knowledge Organization* also include lists. Some clearing houses collect copies of thesauri and they are also available online. To access these online lists, access the following website: <http://www.fit.qut.edu.au/InfoSys/middle/cont voc.htm>

- How much literature should be included in the IOAR system? This will determine the specificity of the indexing terms. More specific terms will have to be used if a large quantity of literature needs to be included.
- In which language should the thesaurus be compiled? (For example in English only, or in English, Afrikaans and Zulu? The latter is called a multilingual thesaurus.)
- Is the IOAR system manual or computerised?
- Who will use the IOAR system and thus the thesaurus (e.g. end-users or information specialists)? This will influence the choice of terms.
- What potential questions will be put to the IOAR system (for which queries will the thesaurus be used to identify the search terms)? When terms are collected for thesaurus entries, they must have literary and user warrant. Will the request be general or specific? If the request is specific, then the terms in the thesaurus must also be specific.

## Step 3: Collect Terms to Include in the Thesaurus

The following sources can be used:

- Existing lists of terms (e.g. dictionaries and other thesauri).
- Subject literature. Terms appearing in the literature will be used by the IOAR system; these are known as terms with literary warrant. This approach is sometimes referred to as the empirical approach.

*Two methods can be used here: the **deductive** and **inductive** methods.*

*With the **deductive** method terms are derived from entities and checked by experts. Broad classes are identified and then other terms are assigned to these classes on the basis of their logical relationships.*

*In the **inductive** method new terms are selected as they appear in entities for possible inclusion in the thesaurus. As each term is recorded, it is classified under a broad class. The thesaurus is constructed right from the beginning to have broader and narrower terms. Experts can be called in to help determine terms.*

*A combination of these methods can also be used.*

- Subject experts. Experts can be approached to compile a list of terms or a committee can be established to decide on terms. These terms are known as terms with user warrant. This process is referred to as the committee approach or the bottom-to-top approach.

## Step 4: Select and Formulate Indexing Terms

The thesaurus compiler(s) must decide which of the terms that have been identified should be included in the thesaurus. A decision must also be made regarding preferred terms and non-preferred terms.

- Verify that the terms selected are found in the entities and keep user preferences in mind (e.g. if a choice must be made between scientific and everyday terms).
- Consult relevant standards and guidelines when formulating terms.

Examples of important standards include (or any later editions):

British Standards Institution. 1987. *British standard guide to establishment and development of monolingual thesauri*. London: BSI (BS 5723).

International Organisation for Standardisation. 1985. *Guidelines for the establishment and development of monolingual thesauri*. ISO 5964. Geneva: ISO.

International Organisation for Standardisation. 1986. *Guidelines for the establishment and development of monolingual thesauri*. ISO 2688. 2nd edition. Geneva: ISO.

National Information Standards Organization. 1994. *Guidelines for the instruction, format and management of monolingual thesauri: an American National Standard*. Bethesda, Maryland.

- Use reference works such as dictionaries, encyclopaedias and other thesauri to standardise terms.
- Decide on the specificity of terms.
- Decide on the degree of precoordination necessary.

## Step 5: Arrange the Terms in Categories so that They Are More or Less Related

Classification systems are particularly useful.

## Step 6: Decide on the Structure of the Thesaurus (e.g. Alphabetical or Hierarchical)

If the thesaurus is to be organised alphabetically, use either a word-by-word or letter-by-letter arrangement:

- Word-by-word arrangement. Each word making up an indexing term is regarded separately. For example:
  - Planet-gazing
  - Planetariums
  - Planetary orbits
  - Planets

- Letter-by-letter arrangement. All spaces between words are ignored, as well as all characters (except for parenthesis), numbers and letters. Terms are filed according to the following sequence: left parenthesis, numbers 0-9, and letters A-Z. For example:  
 Planetariums  
 Planetary orbits  
 Planet-gazing  
 Planets

The thesaurus entries should be arranged in alphabetical order. Terms that are listed in each thesaurus entry should also be sub-arranged in alphabetical order. For example:

#### **COOLERS**

- NT Air coolers  
 Beverage coolers  
 Milk coolers  
 Oil coolers
- RT Air conditioning  
 Air conditioning equipment  
 Compressors

## Step 7: Decide on the Format of the Entries

For example:

#### **INDEXING TERM**

- SN (scope note)  
 UF (used for)  
 BT (broader term)  
 NT (narrower term)  
 RT (related term)

## Step 8: Use a Standard Form or an Appropriate Thesaurus Program to Record Thesaurus Entries

- Decide which terms will be preferred terms and which non-preferred terms.
- Indicate relationships between terms.
- Add scope notes where necessary.

## Step 9: Compile a Draft Thesaurus

Ask users and experts to test it for

- indexing of entities
- retrieving information from the IOAR system

Take note of the problems and revise the thesaurus as necessary.

## Steps 10-14

Step 10: Write an introduction for the thesaurus and explain how it should be used.

Step 11: Edit the thesaurus.

Step 12: Copy the thesaurus and make it available online or in printed form.

Step 13: Send a copy to a clearing house.

Step 14: Maintain the thesaurus and evaluate its effectiveness on a regular basis.

Maintenance can be a time-consuming task. Terms have to be updated, new descriptions have to be added, existing entries have to be adapted and obsolete ones deleted.

## Guidelines for Formulating Thesaurus Entries

The following basic guidelines have been taken from standards for thesauri and textbooks and include the many details associated with compiling a thesaurus.

### Indexing Terms

- Noun and noun phrases (e.g. FEMALE WORKERS) are mostly used as indexing terms.
- Use adjectives as indexing terms in exceptional instances only. For example:  
miniature  
portable

However, such terms should rather be used as compound terms in the form of phrases. For example:

miniature portraits  
portable devices

- Avoid prepositions and verbs wherever possible. For example:  
book evaluation (not evaluation of books)  
communication (not communicate)

### Singular versus Plural Form

- Nouns for tangible objects that can be counted (in the literature these are referred to as 'count' nouns) are used in the plural form, for example dogs, trees, children. The question 'how many?' is applicable to count nouns.
- For non-count nouns use the singular form, for example gold, silver, flour, sugar. The question 'How much?' is applicable to non-count nouns.

Be aware that there are exceptions to this rule, about which you will find more information in the literature.

## Spelling

Standard spelling should be used. If a choice has to be made between English and American spelling, the user's preference should be kept in mind. There should always be a reference (U) from the term with the 'wrong spelling' to the term with the standard or preferred spelling. For example:

Archeology  
 U **ARCHAEOLOGY**  
**ARCHAEOLOGY**  
 UF Archeology

## Punctuation

Punctuation marks should be avoided if possible. Hyphens and apostrophes should not be included, for example

multicultural societies (not multi-cultural societies)  
 nonmetals (not non-metals)  
 postcoordinate indexing (not post-coordinate indexing)

The hyphen should be retained if its exclusion could lead to ambiguity or difficult reading. For example:

draw-off taps  
 n-types  
 parent-teacher relationships

Parentheses should be used if a qualifier for an indexing term is required. (A qualifier explains the meaning of a term and puts it into context so that the term is distinguishable from other terms with the same spelling but with different meanings.) For example:

Letters (alphabet)  
 Letters (correspondence)

## Abbreviations

Try to avoid the use of abbreviations. They should only be used as preferred terms if they are very general and well known. For example:

North Atlantic Treaty Organization  
 U **NATO**  
 United States of America  
 U **USA**

## Compound Terms

According to the *International ISO standard* it is preferable to use single terms for a compound concept, unless the compound concept would be clearer to the user. When

compiling a thesaurus it is often difficult to decide whether or not to use a compound term. The number of compound terms that occur in a thesaurus indicates the *level of precoordination* of the thesaurus. If many compound terms occur, the thesaurus will have a high level of precoordination. Examples of compound terms are

- dried fruit
- municipal engineering
- philosophy of education
- rubber legs
- special libraries

When an indexing term consists of two words, preference should be given to the direct form of entry. For example:

- mechanical engineering (not engineering, mechanical)
- special libraries (not libraries, special)

When the selected terms have been arranged in alphabetical order, indicate their relationship with other terms. Make sure that for each term there is a *reciprocal entry*, for example

#### **LIBRARIES**

NT Special libraries

#### **SPECIAL LIBRARIES**

BT Libraries

## Relationships Application

You have already been introduced to the relationships found in thesaurus entries. The following is a simple explanation of how these relationships and reciprocal entries work in practice.

### Hierarchical Relationships

To determine hierarchical relationships, you should ask

Is ... a type of ...?

If the answer is YES, then is it a narrower term? For example:

**Special libraries** is a **type** of library, therefore

#### **LIBRARIES**

NT Special libraries

In the thesaurus a specific entry is made for special libraries: Libraries is then referred to as the broader term. Under both Libraries and Special libraries a number of relationships can be indicated, for example

**LIBRARIES**

- NT Academic libraries
- Public libraries
- Special libraries
- RT Library buildings
- Library personnel

**SPECIAL LIBRARIES**

- BT Libraries
- RT Actuality services
- Information services

For the terms Academic libraries and Public libraries, similar entries are made, and each time Libraries is used as the broader term.

## Non-hierarchical Relationships

To determine non-hierarchical relationships, two questions must be asked:

- Is ... the same as ...?
- Is ... a type of ...?

If the answer is NO to both questions, then you must think of possible non-hierarchical relationships. Non-hierarchical relationships help the user and indexer to think of other terms which might be useful. If, for example, you are looking for information about **Subject analysis**, think of the following: Would information about **Conceptual analysis** be useful? Would information about **Indexing** be useful? Would information about **Abstracting** be useful? You can continue to determine related terms in this way. Related terms are indicated as follows:

**SUBJECT ANALYSIS**

- RT Abstracting
- Conceptual analysis
- Indexing

A thesaurus entry can also be made for abstracting, with Subject analysis as a related term:

**ABSTRACTING**

- RT Subject analysis

Further relationships can be indicated under both Subject analysis and Abstracting. For example:

**ABSTRACTING**

- RT Abstracts
- Indexing
- Subject analysis

**SUBJECT ANALYSIS**

- RT Abstracting
- Conceptual analysis
- Indexing

Similar entries can be made for Indexing and Conceptual analysis – each time they must refer to Subject analysis as a related term.

## Equivalent Relationships

To determine equivalent relationships, one should ask

- Is ... the same as or almost the same as ...?
- Do I want to use ... instead of ...?

If the answer is YES to any of these questions, then you must think of possible equivalent relationships. Equivalent relationships are used to guide the user of the thesaurus (indexer or enquirer) away from the terms not regarded by the system as suitable for indexing, to the preferred terms with which the concept can be described. To indicate these relationships, cross-references are used (U, UF). For example:

<p>Librarianship U <b>LIBRARY SCIENCE</b></p> <p>Library employees U <b>LIBRARY PERSONNEL</b></p> <p><b>LIBRARY PERSONNEL</b> UF Library employees</p> <p><b>LIBRARY SCIENCE</b> UF Librarianship</p>
---

Draw up the thesaurus in typed form – use an existing thesaurus as an example. Keep the thesaurus up to date by making changes, adding new terms and dropping or replacing obsolete terms.

## Thesaurus Software

Take note of available thesaurus software and the advantages this software offers. If you would like to experiment with such programs, you can go to the following websites. Some of them offer demonstration programs that you can download:

LEXICO: <http://www.pmei.com/lexico.htm>

MULTITES: <http://www.multites.com>

QUESTANS LIMITED: <http://www.questans.co.uk>

The Hierarch Thesaurus Manager: <http://www.ozemail.com/au%7Esisnsw/hierarch.htm>

There are special programs available for constructing thesauri. These programs make provision for the following:

- typography and sorting (e.g. correct use of capital and small letters and also special characters)

- displaying the record for each entry (this includes the relationships between terms)
- distinction between non-preferred and recommended terms

When evaluating such a program for possible use, ensure that the program

- does not limit the number of characters in a term, the levels of hierarchy, or the number of relationships that may be indicated
- is able to handle searches on full terms (e.g. libraries) and also truncated terms (e.g. librar...)
- shows entries in such a way that a term can be displayed with all its relationships
- makes it easy to create and edit terms
- supports the tracing of errors (e.g. duplicate entries and inconsistent application of relationships)
- makes cross-references (relationships should be correctly indicated and terms to which there are cross-references should actually exist)
- is able to delete terms
- is able to add candidate terms provisionally
- allows other fields (e.g. classification notations) to be included
- provides printed reports from the thesaurus in a format required by the user
- is able to copy a sub-collection or a micro-thesaurus of thesaurus entries from one thesaurus to another

## BIBLIOGRAPHY

- Asis thesaurus of information science and librarianship*. 1994. Medford, NJ: Learned Information.
- Batty, CD. 1989. Thesaurus construction and maintenance survival kit. *Database*, 12(1):13-20.
- Aitchison, J, Gilchrist, A & Bawden, D. 1997. *Thesaurus construction and use: a practical manual*. 3rd ed. London: Aslib.
- Borko, H & Bernier, CL. 1978. *Indexing concepts and methods*. New York: Academic Press.
- Lancaster, FW. 1986. *Vocabulary control for information retrieval*. 2nd ed. Arlington, Va: Information Sources Press.
- Lancaster, FW & Warner, A. 1993. *Information retrieval today*. Arlington, Va: Information Resources.
- National Information Standards Organization. 1994. *Guidelines for the instruction, format and management of monolingual thesauri: an American National Standard*. Bethesda, Maryland.
- Orna, E. 1983. *Build yourself a thesaurus: a step-by-step guide*. Norwich: Running Angel.
- Ovens, CS. 1989. Methodology for constructing a thesaurus in the linguistics and literature fields. *South African Journal of Library and Information Science*, 57(1):45-50.
- Taylor, AG. 2004. *The organization of information*. 2nd ed. Westport, Conn: Libraries Unlimited.
- Thesaurus of ERIC descriptors*. 1995. Edited by JE Houston. 13th ed. Phoenix, Ariz: Oryx Press.
- Walker, G & Janes, J. 1998. *Online retrieval: a dialogue of theory and practice*. Englewood, Colo: Libraries Unlimited.



Umashanie Reddy and Omesh Jagarnath

### Abstract

An index is the soul of a book, and if a book is worth publishing at all, it is worth an index (Hewitt 2003:201). According to Cleveland and Cleveland (1990:100) good indexing is not a casual job but the result of a professional activity carried out by people with proper training and experience. Procedures and techniques acquired over the years have to be learnt and followed. The purpose of this chapter is to familiarise the reader with the technical presentation and editing procedures of a book index. Many of the issues in this chapter have been covered in other chapters as well, but since this is such an important part of the indexing process, this chapter provides further examples and explanations. The general principles can also be applied to periodical indexes and thesauri.

### Introduction

The editing process of indexes cannot be clearly distinguished from the selection and formulation of entries. Most indexers edit briefly as they work, making sure that each new entry is free of errors and that it conforms to existing entries. Others prefer to input all their raw entries as fast as possible without any checking or revision. On completion, the index is scrutinised in its entirety to standardise the headings, add reversed or permuted headings, construct subheadings, add cross-references, ensure consistency of style, and the like. Others prefer to proceed more slowly and control the structure of the growing index. For example, when subheadings become necessary (due to the magnitude of locators, or because several different aspects of the subject are identified) the locators are grouped accordingly by appending suitable words or phrases; anomalies in names and variations in spelling and capitalisation are checked when they arise; alternative headings (reversed or permuted) and cross-references are created when the need arises. *No kind of preliminary editing cancels out the need for final editing of the index.* It is important to remember that perfection is often not achievable in the formulation of index entries – to observe consistency in one feature can cause inconsistency with regard to another, and to follow a technically correct guideline can result in something that looks odd in a particular index. A pragmatic approach, treating the text with respect and making access as simple as possible for the user, is the best (Booth 2001:127-128).

## Editing and Presentation

The finished index must provide a balanced and relevant coverage of the entity's significant content; it must enable access by all likely terms, be clearly laid out, be preceded by an introductory note (if necessary) and meet the specifications agreed upon by the publisher. *Editing ensures consistency of form and structure. Presentation influences the readability and use of the index.* Users must find the terms they are looking for with ease and understand them unambiguously, as well as note relevant references. Presentational features are normally already in the indexer's mind during input and some can be incorporated from the start, for example capitalisation, punctuation, indication of major references and illustrations. The final presentation, spacing and layout, additional styling and preparation for the chosen form of copy (paper, disk, camera-ready) can only be given attention once the individual entries and the overall structure is in place (Booth 2001:130).

Whether the index is edited on screen, or on a print-out from which the corrections are then keyed in, depends on the indexer's preferred method of working. Working on a print-out takes extra time, but it is easier to scrutinise a print-out, page back and forth to compare entries, see where improvements and corrections are needed, make notes and keep track of amendments. The last opportunity for editing an index for a printed entity is during the checking of the index page proofs.

A poor index is difficult to repair without an extensive makeover. At this stage the focus should be on such things as misspelling, large blocks of undifferentiated locators, too many subheadings under a heading, related information scattered throughout the index without cross-references. Dead-end cross-references should be corrected and page reference accuracy should be randomly checked. According to Cleveland and Cleveland (2001:190-191), the tasks can be summarised as follows:

- Correct alphabetising.
- Divide long multi-level headings.
- Eliminate synonymous headings by consolidation under preferred terms.
- Recheck the editing.
- Correct too many locators attached to an entry.
- Correct ambiguous headings.
- Verify cross-references.
- Correct spelling and capitalisation.
- Check punctuation and capitalisation.
- Check need to add entries.
- Ask: Are main headings relevant and/or needed?
- Ask: Are locators correct?
- Ask: Is the index overdone?
- Ask: Is the index underdone?
- Bottom line: Does the index work?

## Factors Influencing Editing

Booth (2001:132-135) explains 11 factors influencing the amount of editing to be done. In this chapter, these are only mentioned briefly since they may differ from one index/indexer to another, but should be attended to:

- Quality of the input. As already mentioned, it depends on the indexer's preference when to edit (while indexing or after the index is completed). If more time is spent during the input stage, less time will be required for editing at the end.
- Length and complexity of entries. An index containing many multiple-word entries, or entries with several subheadings or sub-subheadings, or having complicated locators, will need more editing time.
- Number of problems noted during input. Indexers often note errors and variations in spelling in the text. Rather than spending time on each one as they arise by consulting the author, editor or publisher, it is better to deal with them together during editing.
- Incidence of special characters in entries. These include mathematical symbols, accented letters, other alphabets, and so on, and may not be available on the indexer's keyboard; or if available, they may be reproduced differently when transferred to another system (e.g. at the publisher). Special characters not available on the keyboard can be marked during input by using other characters so that they can be converted or noted during editing. This will need more editing time.
- Number of entries. The longer the index, the longer it takes to edit, and the more difficult it is to spot inconsistencies between entries.
- Mismatch between the length of the index and the space available for it. Even though the indexer knows from the start how much space is available for the index, it may still be too long. If it is not possible to make more space available by using a smaller font than originally agreed upon, the indexer is faced with the difficult and time-consuming task of reducing or removing entries. It may also happen that the index is too short for the allocated space.
- Style requirements. Most features like capitalisation, punctuation, use of italic or bold typeface can be done during the input stage. Others must be left until the final editing, for example putting locators for major references in bold type.
- Degree of autonomy of the indexer. Usually the indexer is entirely responsible for the content of the index – the publisher or editor only ensures that the submitted index meets the agreed specifications. Sometimes the index needs to be approved by the editor, publisher or author of the text. If, for example, suggestions are made for the addition and/or removal of entries, or changes of wording, the index should be referred back to the indexer.
- Time available. It can happen that the deadline for the index does not allow for thorough editing. In such cases it is better to edit the index while working on it. The indexer can try to negotiate for more time, but it is not often granted.
- Producing the index. The physical form in which the index is prepared varies according to the requirements of the publisher: e-mail, on disk, hard copy, et cetera. A hard copy

gives a clear view of the index and its layout and allows for notes to the publisher regarding special features. Be aware that the page dimensions, the font and size used in a typeset document are almost always different from those used by the indexer. Pay particular attention to subheading sequences that may be broken by ends of columns or pages. This is especially important when a camera-ready copy is required (in instances where the indexer has the suitable software). In-house writers producing documents on desktop publishing systems may compile their own indexes using embedded codes placed in the text. For more information on this, see chapter 8 on embedded indexing.

- **Proofreading.** This takes place after the index has been printed. It should be done by the indexer, but is often dealt with by the editorial staff of the publisher.

At all times the indexer should ensure familiarity with the agreement and adhere to its stipulations regarding the technical presentation of the index. Should problems arise, contact the editor or publisher and discuss the alternatives. The following paragraphs cover specific issues – some issues can be dealt with while constructing the index, while others should be left until the final editing.

## Specific Tasks

### Check for Errors

Activate a spell check during the indexing process. Typing and spelling errors will thus be indicated. If indexing on cards, each entry must be checked for correct spelling. Ideally this should be done by another person (time permitting). It is imperative in book or periodical indexing to adhere to the terminology used by the author. One's own preferences cannot substitute those of the author, for example 'archeology' for 'archaeology' or 'organization' for 'organisation'. When compiling a thesaurus, system preferences will have to be followed. Some indexers prefer to do a character-by-character (instead of a word-by-word) reading of the index. With a word-by-word reading the visual impression of some indexing terms might 'look' correct, for example 'geranium' instead of 'germanium' or 'Van der Berg' instead of 'Van den Berg'. If the numeral '0' and the letter 'O' are incorrectly keyed in, a computer-sorted index will misplace the entry.

### Check for Scattering

Make sure that a single topic is not scattered under different word forms, synonyms, or near-synonym words in the index, for example 'goose' and 'geese', 'dexterity' and 'dexterous', or 'civil law' and 'common law'. If terms are left scattered under more than one heading, it may confuse users. Rather use just one form for each topic (e.g. plural, nouns). Some instances may require cross-references.

### Check Multi-word Headings

These are also known as compound terms or complex terms, and can also be phrases, that is consisting of more than one word. Some examples are 'descriptive cataloguing',

'computer-assisted indexing' and 'progressive subheading order'. Check such headings for possible additional entries, for example 'cataloguing', 'indexing' and 'subheading order'. Under these broader terms the specific terms may be listed, or cross-references may be used, depending on the structure of the index. At this stage, also check too long headings – they can be reduced by the removal of prepositions for example, or perhaps rephrased.

## Check Cross-references

Cross-references can be added while indexing, but at the final editing they should be checked in the context of the full index. Some may be discarded or others may be added. All cross-references must lead to headings that exist in the index and that are followed by locators, otherwise dead-end cross-references will ensue. Cross-references should be used only when necessary. Some headings may require more than one cross-reference. Examples of cross-references:

<p>cataloguing <i>see also</i> subject cataloguing  dedicated indexing software 207-209  embedded indexing 38, 121-127, 206    editing 130-133    software 210-215  indexing    as classification 15-19    defined 5-6    policies 22-32, 45  indexing software <i>see</i> dedicated indexing software    embedded indexing 210-215  subject cataloguing 15-19  subject indexing <i>see</i> indexing</p>
--

## Check Locators

The accuracy of locators (e.g. page numbers, volume and part numbers, columns) is extremely important since it affects the access to the information contained in the entity. Errors (including possible typing errors) should be constantly checked to ensure accuracy. Page ranges must be consistent, for example 331-337, or 331-37, or 331-7 (full ranges are often preferred to reduced forms). Long strings of locators (i.e. without subheadings to indicate parts or points) are mostly irritating to the user since he or she must work through all locators to find a needed piece of information.

For example:

metadata 14, 69, 91-92fig, 123-125, 139, 141, 143-144, 147-148,149, 151-152, 153-154, 155, 159, 164, 190-192, 202, 203, 335-341

There are different ways of solving this problem, of which only the most frequently used two are illustrated. Bold type can be used to highlight locators leading to major information or emphasising a particular point:

metadata 14, 69, 91-92fig, 123-125, 139, 141, 143-144, 147-148,149, **151-152**, 153-154, 155, 159, 164, 190-192, 202, 203, **335-341**

A more sensible solution is to use subheadings:

#### metadata

- access to 149, 202
- administrative 147-148
- arrangement 335-341
- characteristics 143-144
- content 69
- creation 159, 190-192
- definition 139
- design 143
- granularity 141
- implementation 151-152
- maintenance 91-92fig
- mapping between formats 155
- preservation 148-149
- registries 153-154
- searching 123-125
- standards for 164
- structure 142, 150-151
- systems 140
- technical 150-151
- types of 141, 152, 203
- use of 14

All the locators should be represented by subheadings – there must be no leftover locators. However, sometimes long strings of locators are unavoidable, for example the listing of locators where an author is cited in the text of an entity:

Lancaster, FW 3, 15, 26, 34, 40, 61, 68-69, 79, 84, 89, 92, 101, 113, 126, 148, 184, 188-189, 236-237, 239, 244-245, 292, 296, 320-321, 383-384, 396

## Check for Qualifiers

Some headings may need a qualifier to make its meaning clear, for example when a word has more than one meaning or if persons share the same name:

- navigation (aeronautics)
- navigation (astronautics)
- Barnard, Chris (author)
- Barnard, Chris (surgeon)

## Check Headings Beginning with the Same Word

Headings beginning with the same word should be arranged in such a way that there is no unwanted separation of terms or odd-looking subheadings. Therefore:

mass-energy equivalence  
 Mass, Jason  
 mass media  
 mass media music  
 mass (substance)  
 Mass Transfer Union  
 mass transportation

is preferable to

Mass  
 -energy equivalence  
 Jason  
 media  
 media music  
 (substance)  
 Transfer Union  
 transportation

## Check Alphabetical Arrangement

Preference must be given either to word-by-word arrangement or letter-by-letter arrangement; subheadings must follow the same arrangement as the main headings; take into account punctuation, prepositions, articles, et cetera; look at chronological sequences; hierarchies, progressive arrangement, and the like.

### Word-by-word and Letter-by-letter Arrangement

In the word-by-word method, spaces between words have value so that shorter words come before longer ones, for example 'snake charmers' will be filed before 'snakebites'. Hyphens (and dashes and diagonal back or forward slashes) are treated the same way as spaces. For example:

snake charmers  
 snake dances  
 snake-eels  
 snake-flies  
 snake plants  
 Snake River  
 snake venom  
 snakebites  
 snakeheads (Fishes)  
 snakes

In the letter-by-letter method, spaces and other kinds of characters are ignored. For example:

snakebites  
 snake charmers  
 snake dances  
 snake-eels  
 snake-flies  
 snakeheads (Fishes)  
 snake plants  
 Snake River  
 snakes  
 snake venom

The other kinds of arrangement mentioned at the beginning of this section (e.g. hierarchical, chronological, progressive) are not discussed since they appear only in certain types of index, for example in indexes to history books, or in manuals where technical processes or developmental stages are discussed. Also pay attention to initial articles, abbreviations, prefixes, numerals, broken order, diacritics and special characters.

## Layout of Entries

Subheadings should be clearly distinguishable from the headings to which they are subordinate, as well as from any sub-subheadings that may follow. Two methods are used: setting out and running on. The set-out method gives a clearer display and can accommodate many levels of subheading, while the run-on method takes up less space, but is limited in the number of subheading levels it can handle. In the set-out layout, each subheading (wherever there is more than one) starts on a new line, with indentation used to show subordination – any further levels of subheading are progressively and evenly indented. In the run-on layout subheadings continue on the same line, with punctuation indicating subordination and when the line is full, it turns over to the next line that is indented by one space. Only two levels of subheading can be accommodated if the subheadings are run on directly after the heading. A combination of the two methods is also possible (Booth 2001; Mulvany 1994).

The following examples illustrate the two layouts:

### *Run-on layout:*

Internet: communication 10-18; downloading 317-319; networks 5; search engines 215-217 (*see also* searches); tools 20

### *Set-out layout:*

Internet  
     communication 10-18  
     downloading 317-319  
     networks 5  
     search engines 215-217 *see also* searches  
     tools 20

## Spaces and Typestyle

Spaces between the letters of the alphabet in the index makes it clear where a new letter starts and ease the use of the index. For example:

*Sears list of subject headings*  
 subject analysis  
 systems

templates  
 terminology

Unger, Richard  
 Unicode  
 user behaviour

If space is available, a letter can be placed at the beginning of each letter sequence. For example:

**S**  
*Sears list of subject headings*  
 subject analysis  
 systems

**T**  
 templates  
 terminology

**U**  
 Unger, Richard  
 Unicode  
 user behaviour

Please note that it is common practice to present titles of works and terms in foreign languages (such as Latin and Greek) in italics. Also ensure that the use of punctuation between page numbers and the use of brackets for qualifiers is consistent.

## Conclusion

This chapter has outlined basic principles on the technical features of presenting a book index. However, some issues were not covered in this chapter. The best source to consult for guidelines is a standard such as the BS ISO 999 (see the bibliography that follows) and/or a reputable indexing manual (see the bibliography). Should the appearance of the published version of the index be unsatisfactory when received at the page-proof stage, the publisher should be consulted.

## BIBLIOGRAPHY

- Booth, PF. 2001. *Indexing: the manual of good practice*. München: Saur.
- Borko, H & Bernier, CL. 1978. *Indexing concepts and methods*. New York: Academic Press.
- British Standards Institution. 1996. *Information and documentation – guidelines for the content, organization and presentation of indexes*. BS ISO 999:1996. London: British Standards Institution.
- Butcher, J. 1992. *Copy-editing: the Cambridge handbook for editors, authors and publishers*. 3rd ed. Cambridge: Cambridge University Press.
- Cleveland, DB & Cleveland, AD 1990. *Introduction to indexing and abstracting*. 2nd ed. Englewood, Colo: Libraries Unlimited.
- Cleveland, DB & Cleveland, AD. 2001. *Introduction to indexing and abstracting*. 3rd ed. Englewood, Colo: Libraries Unlimited.
- Gregory, GM. 2000. Current trends in book indexing. *Information Today*, 17(5):69.
- Hewitt, AR. 2003. Another 'How to make an index'. *The Indexer*, 23(4):200-201.
- Lancaster, FW. 1998. *Indexing and abstracting in theory and practice*. 2nd ed. Champaign, Ill: University of Illinois, Graduate School of Library and Information Science.
- Lipetz, B-A. 1989. The usefulness of indexes, in *Indexing: The state of our knowledge and the state of our ignorance: Proceedings of the 20th Annual Meeting of the American Society of Indexers*, edited by BH Weinberg. Medford, NJ: Learned Information:111-120.
- Mulvany, NC. 1994. *Indexing books*. Chicago: University of Chicago Press.
- University of Chicago Press. 1993. *The Chicago manual of style*. 14th ed. Chicago: University of Chicago Press.
- Wellisch, H. 1995. *Indexing from A to Z*. 2nd ed. New York: Wilson.