
Module A: Sample

Aims and objectives ♦ Contents ♦ Text extract

Training in indexing: 4th edition
(Module A revised 2014)

© Society of Indexers 2010 –14

Module A

Study guidelines

- This module is a general introduction to indexing. You will find more detailed coverage of all aspects of indexing in the later modules.
- Some of the ideas peculiar to indexing are necessarily described at length. It is important to remember that indexing requires knowledge (general and special), a mixture of common sense and reason, plus an element of imaginative helpfulness – all combining to help the index user.
- While studying the module, examine in detail as many indexes (of different types) as you can. Check whether the entries are accurate and useful; whether there are obvious omissions; whether there are helpful cross-references; in what order the entries are arranged. Have a couple of indexed books beside you while studying the module, so that you can find examples of what you are reading about. You should also practise compiling indexes to anything you read.
- Remember that there are a great many additional resources on the training course website; make full use of them.
- Before sitting the test paper for this module, work through the practice exercises for this module on the training course website to check whether you are ready for formal assessment.

Aims and objectives

Aims

The general aims of this module are to:

- define some basic terms used in indexing
 - describe what indexes are for
 - identify the main characteristics of indexes
 - identify what users want from indexes
 - identify what kind of people make good indexers
 - describe how indexes are made
-

- describe the role of authors and other document originators
- describe how various types of document are produced.
- describe why human analytical indexing provides a superior tool to full text search or to 'automated indexing'.

Objectives

After studying the module, students should be able to:

- understand the terms used in indexing and publishing
- describe how indexing fits into the document production process
- understand the importance of indexes
- understand the needs of index users
- identify the knowledge and skills that indexers need
- describe how an index is made
- understand the aims and intentions of authors and other originators
- understand how documents are produced
- understand the conventions of bibliographic references
- identify different types of document
- name the basic reference sources used by indexers
- create a short and simple standalone index.
- argue the case for human analytical indexing
- argue the case for indexes in eBooks

Contents

Module A: Introducing Indexing: indexers, users and documents	
A1 Setting the Scene	1
A1.1 Human analytical indexing	2
A1.2 Definitions	3
A2 Index functions and characteristics	15
A2.1 Function	15
A2.2 Necessary characteristics	16
A2.2.1 Main headings and subheadings	16
A2.2.2 See <i>also</i> cross-references	17
A2.2.3 See cross-references	17
A2.2.3.1 Double entry	17
A.2.2.4 Locators	18
A2.2.4.1 Locators in tagged indexes	20
A2.2.4.2 Locators in embedded indexes	20
A2.2.4.3 Locators in eBooks	20
A2.3 Common practice: conventional usage	23
A2.4 Indexable elements: what to put in the index	26
A2.4.1 Thesauri and classification schemes	29
A2.5 Direct entry and classified indexing	29
A2.6 The position of the index	30
A2.7 Single and multiple indexes	30
A2.8 Summary: types of index entry and arrangement	31
A3 Users of documents and indexes	33
A3.1 The purpose and use of documents	33
A3.2 Users of documents	35
A3.3 What users need from an index	35
A3.4 Summary: the needs of the index user	36
A4 Indexers	37
A4.1 Situations requiring an indexer	37
A4.2 Who is the indexer?	40
A4.3 Knowledge and skills	41
A4.3.1 Reference sources for the indexer	42
A4.3.1.1 Online reference sources	42
A4.3.1.2 Library reference sources	44
A4.3.2 The indexer's personal library	44
A4.4 Indexing equipment	45
A4.5 The indexer's health	46
A4.6 Summary: the professional indexer	47

A5 The index	48
A5.1 Index coverage, size and detail	48
A5.2 Index components	49
A5.3 Index titles	50
A5.4 Making the index	50
A5.5 Computers in indexing	54
A5.5.1 Indexing software	55
A5.6 Summary: compiling the index	56
A6 Authors and other originators	57
A6.1 Authors and their intentions	57
A6.2 Types of authors and originators	58
A6.2.1 Multiple authors	58
A6.2.2 Creators of related documents	59
A6.2.3 Anonymous originators	59
A6.2.4 Identifying originators	59
A6.3 Summary: indexers and authors	60
A7 Documents and the production process	61
A7.1 Document creation	61
A7.1.1 Editing and proof correction	61
A7.1.2 Publication	62
A7.1.3 Document identification	63
A7.1.3.1 Bibliographic citation	63
A7.1.4 Access to contents	66
A7.1.5 The document/index relationship	66
A7.1.6 Document modifications	67
A7.2 Types of documents	68
A7.2.1 Physical format	68
A7.2.2 Method of presentation and arrangement	68
A7.2.2.1 Abstracts	69
A7.2.3 Purpose and users	70
A7.2.4 Language	70
A7.2.5 Edition size	70
A7.2.6 Primary and secondary sources	71
A7.2.7 Factual and non-factual status	71
A7.2.8 Currency	72
A7.3 Summary: document production and description	72
A8 Memoranda for indexers	74



Setting the Scene

Introduction

Traditionally an index has been constructed as an A-Z list of concepts and topics contained in a document, providing both a map of the text and a retrieval tool for information. Good indexing principles and techniques have been developed over time and these form the basis of current good practice in indexing and make up the main content of the course. The basic principles for producing a well-structured index are relevant for both traditional printed indexes and electronic forms of publication, as the need to analyse a text to provide intelligent access to it is an intellectual process.

The role of the indexer

Indexers use their knowledge of the subject and their skills in providing access to information to create a structured list of named and unnamed (implied) concepts covered in the text, drawing scattered references together and analysing larger topics into their component parts. The indexer will think empathetically about potential alternative approaches to finding information that the user may make (e.g. searching under **synonyms**) and provide appropriate cross-references to accommodate these. In addition the indexer will point the reader to other related discussions that may be of interest to the user who has looked up a particular term. These are not decisions that can be made by automated and semi-automated indexing programs, which are based on word-spotting. An intelligently crafted index will be a useful, easily navigable tool for the users, leading them to the information they need by the shortest, quickest route.

The digital revolution in publishing

eBooks are a new delivery format. They are now being produced alongside printed books with the whole publication often '**born digital**' i.e. produced in a digital format from the start, allowing export to a variety of platforms including print and eBook formats. While eBooks also provide for full-text searching as a means of accessing information,

an index based on an intelligent analysis of the text remains a useful tool which can be used on its own or in association with other search facilities.

Traditional printed indexes are constructed using fixed locators as reference points (**page numbers**, paragraph numbers etc.). An eBook by contrast has '**reflowable text**' – the text size can be altered according to the user's preference and the book is then 'reflowed'. Depending on the print size chosen the book will take up more or fewer screen 'pages'. Pages as fixed points of reference therefore no longer exist. This means that they cannot be used as locators for index entries; instead the index entry has to be an active link to a specific anchor embedded in the text that will remain associated with the specific piece of information as the text is reflowed.

There are various ways of doing this; some publishers use **XML**-based **workflows** in the production process which allow use of embedded indexes (indexes generated from entries embedded in the text) or tagged indexes (indexes supplied separately with entries linked to identifiers in the text).

One of the advantages, for the user, of active linking is that references lead much more specifically to the information – indexing to the sentence or word level (as opposed to the printed page) is possible and is known as **pinpoint indexing**.

A1.1 Human analytical indexing

So what are the advantages of human analytical indexes over indexes produced by automated and semi-automated indexing programs or using a simple text search?

An indexer will

- provide alternative access to terms for users who search under synonyms by providing **see cross-references** (or double entry)
e.g. pupils see students
- recognize relationships with other topics, and provide appropriate **see also cross-references** to alert the user to these **related terms** e.g. literacy see *also* reading
- make an intelligent analysis of a text so that readers interested in (for example) globalization are led to relevant discussions on international trade where globalization is not actually named
- distinguish between **homographs** (words that look the same but have different meanings) so that for example, readers interested in the pop star Madonna are not directed to material on religious iconography
e.g. Madonna (entertainer)
 Madonna (mother of Jesus)

- provide context for references e.g. through **modifiers** or **subheadings** for groups of **locators** and **page ranges**; or by typographical distinction of locators
- distinguish between important information and very minor or irrelevant references
- pick up references to variant word forms, where a simple text search would only return the word variation searched
e.g. mouse/mice; defences/defensive positions
- highlight visual information contained in images or other graphics.

For a more detailed understanding see

Human or computer produced indexes? by James Lamb

<http://www.indexers.org.uk/index.php?id=463>

Indexers vs keywords by Bill Johncocks

<http://www.ptg-indexers.org.uk/members/locked/keywords.htm>

(PTG website member log-in required: details on main SI website Members' homepage/Resources tab)

It is important to fully comprehend the arguments above for the value provided by a professional analytical index, as they may feature in future discussions with clients and influence your ability to market yourself as an indexer.

There is no substitute for intelligent analytical indexing by a human indexer who has analysed the text in detail, identified the concepts discussed, recognized the inherent relationships between topics, anticipated the alternative approaches that might be made by users, provided pertinent entries with helpful modifiers and subheadings, and organized the whole into an efficacious arrangement so that the user can easily scan entries to find what they need.

A1.2 Definitions

Indexers should be familiar with the following terms, which are standard usage in the indexing and publishing world, so that they are able to use them correctly when speaking or writing to publishers and their editors. In practice, there is some variation in usage and precise meanings. The definitions given here are intended only to inform the student indexer. Fuller discussion of many of these terms will be found in the relevant parts of the training course. Terms in bold are included in this list of definitions. All terms included in this list are also highlighted in **bold** at first mention in the modules.

N.B. The Society of Indexers Publishing Technology Group (PTG) website has a more extensive glossary related to **eBook indexing** which should be consulted for a fuller understanding www.ptg-indexers.org.uk. Terminology is still developing in relation to eBooks, so you should be proactive in keeping up-to-date.

- abridged edition**—a shortened version of a **document** (often intended to simplify a text), e.g. an edition of an adult novel for children or, for adults, editions such as Reader's Digest condensed books
- abstract**—a summary of the information contained in a **document**, e.g. a 100-word precis of a **periodical**/journal article; often collected in journal form, e.g. *Chemical Abstracts*
- alphabetization**—the process of sorting index **entries** into alphabetical order, according to a set of rules; see *also* **letter-by-letter**; **word-by-word**
- antonym**—a term with the opposite meaning to another term, e.g. 'optimism'/'pessimism'
- app**—short for application software and most commonly used in relation to small programs used on mobile devices
- archives**—**documents** no longer used for their original purposes but preserved for research, e.g. historical papers in the National Archives, or redundant institutional records
- audiovisual document**—a **document** in which the information is presented as images and sounds, e.g. film, videotape, **CD-ROM**
- author**—a person or organization responsible for producing all or some of the information content of a **document**
- back-of-the-book index**—an **index** to the content of a book or similar **document** (pamphlet, brochure, etc.), produced as an integral part of the published document
- bibliographic citation/bibliographic reference**—a precise description of a **document** or part of a document, providing a unique identification
- bibliography**—a list of documents having a common characteristic, such as authorship, place of publication, or subject, e.g. a list at the end of a book or **periodical**/journal article, or a complete volume listing all an author's works, naming all known **editions** and translations
- book packager**—working for a **publisher** who outsources the work, the packager executes various stages of publishing using their own staff or freelance professionals. A packager will therefore work for more than one publisher
- born digital**—**document** created in a digital or electronic form as opposed to a being created through scanning or reformatting original printed material
- camera-ready copy**—an index prepared for production without a **typesetter**; the index in **hard copy** is exactly as it will appear in the publication
- catalogue**—a list of the **documents** or artefacts held in a given collection (e.g. an exhibition, library, archive) or published by a given organization, such as an art gallery or a **publisher**

CD-ROM (Compact Disc Read-Only Memory)—a compact disk containing pre-recorded digital data that cannot generally be amended by the user, e.g. bibliographic information

circular cross-references—**cross-references** leading only to each other or to another **cross-reference** and not to any **locators**

citation index—an **index** to cited **authors** and **documents**, based on the principle that researchers are also likely to be interested in articles citing (referring to) the article they are currently reading, e.g. *Science Citation Index*. This type of **index** is popular with, for example, lawyers and scientists.

classification scheme—a presentation of relationships between subjects structured to facilitate indexing and information retrieval, in libraries, on websites, in current awareness publications and in any environment where large amounts of information must be organized. The notation (coding) of a classification system usually expresses the relationship between subjects. A good classification can be invisible to the user, but it may be used for signposting as in a library, such as 330 economics, 570 biology, 820 English literature (the numbers are taken from the Dewey Decimal Classification).

classified index—an **index** in which the entries are arranged in a **hierarchical structure** of main headings, subheadings and sub-subheadings, rather like a **classification scheme**; see also **direct entry index**

computer medium—where a computer-generated and computer-held record is held, e.g. a hard disk, **CD-ROM**, flash drive or memory stick

concordance—an alphabetical list of the principal words of a **document** or a collection of documents. The most famous concordances are those to the Bible and to the works of Shakespeare.

continued statement—a heading repeated at the top of a page where a page turn interrupts the entries following a main heading, e.g.

Dickens, Charles (*cont.*)
public readings 57

copy—material to be printed; see also **hard copy**; **manuscript**

cross-reference—an instruction directing the user from one **heading** or **subheading** to another, e.g.

movies *see* films
dairy products *see also* cheeses

Locators are not included in **cross-references**.

cumulative index—an **index** to a run of issues of a **periodical**

database—an organized digital collection of retrievable material. A bibliographic database encompasses a wide range of formats including journals, conference proceedings, reports, government publications and books.

digital publishing—see **eBooks**; **ePublications**

direct entry index—an **index** in which specific terms are used as main **entries**, in contrast to a **classified index**

document—any item in which information is held, e.g. a book, a computer disk, an **eBook**, a film, a Greek vase, a map, a painting, a slide, a sound recording, a video presentation, a web page, etc.

double entry (double posting)—instead of using a **see cross-reference**, it may be more helpful to the user to make two entries for the same term or synonym, e.g.

engineering, mechanical 22, 38

mechanical engineering 22, 38

eBook—a form of electronic publication, usually corresponding to a printed book but downloadable for display on a handheld device (a dedicated eReader, smartphone or tablet). eBooks are encoded in a form of **HTML** and are similar to a long webpage. There are no fixed **pages**, and the text size can be changed, thus altering how much text is displayed on the screen page. This means that there are no **page numbers** to use as index locators and instead index entries need to be hyperlinks to anchors placed in the text.

eBook index—a digital index where the index term is hyperlinked to the relevant text location: clicking/touching the index term or an accompanying locator symbol will display the appropriate section of text on the reader's screen.

See also the presentation on '*Understanding eBooks*' in the Resource Centre of the SI indexer-training website

edition—a version of a **document**, all copies of which are in the same physical form (binding) and printed from the same (or almost the same) master copy or type image; one edition may run to several printings or **impressions**. A new or **revised edition** will include a substantial amount of new material and/or corrections and is likely to need re-indexing.

editor—a person responsible for the overall preparation of a **document** by either a single **author**, or (more commonly) of a **document** with contributions by different authors; 2. in a publishing house, a person responsible for supervising the content, scope and structure of a **document**. A wide variety of names are used for the editor who is likely to be an indexer's possible client, e.g. production editor, desk editor, managing editor. A copy-editor is responsible for preparing an author's manuscript for publication by, for example, correcting inconsistencies, errors and ambiguities and ensuring that it conforms to house style. A commissioning editor normally commissions an author to write a particular **document** and in some circumstances may also be an indexer's client.

elision (minimal numbering)—the contraction of **locators** referring to a **page range** to produce a shorter form, e.g. 45–48 becomes 45–8; 191–196 becomes 191–6

em—a unit of typographical measurement, derived from the width of the letter ‘m’. An em-rule (or dash —) is twice the length of an **en**-rule and may sometimes be used to separate a word or phrase from the rest of the sentence. **Indentation** is normally measured in terms of ems.

embedded indexing—a technique in which the actual index entries are placed at the point in the text that they relate to: the software embeds the entry in the underlying code of the **document**. The index is generated by the software on demand, with **page numbers** produced to match the current **pagination** of the document. There is no separate index file. Conventional indexing is sometimes described as **standalone indexing** to contrast it with embedding. Embedded indexes can be used by publishers working with **XML workflows** to generate linked indexes in **eBooks**.

en—a unit of typographical measurement, half the width of an **em** and derived from the width of the letter ‘n’. An en-rule is a short dash normally used instead of a hyphen in **page ranges** (e.g. 3–4, rather than 3-4), in phrases involving comparisons (e.g. ‘cost–benefit analysis’) and to separate a word or phrase from the rest of the sentence.

end matter—material normally coming after the main text of a **document**, e.g. end notes, bibliography, maps, **index**

entry—a record in an **index** consisting of a heading followed by a **locator(s)** or a **cross-reference**; see *also* **subentry**

EPUB—an **XHTML**-based standard for electronic publishing, used for tablet eBook readers but not by the Amazon Kindle

ePublications—a blanket term covering all electronically delivered publications formatted and accessible as individual titles equivalent to hard copy books. The two main classes are **eBooks** (viewable on a handheld eReader like the Amazon Kindle or Apple’s iPad) and online paginated books in **PDF**.

exhaustivity—the extent to which concepts and topics in a **document** are assigned **entries**; a function of what the indexer chooses to **index** and the space available; see *also* **specificity**

filing order—sequence in which **entries** are arranged, e.g. alphabetical order

folio—a sheet (i.e. **page**) of an **author’s manuscript**

font—the typeface in which a **document** is printed

frame—1. an area of the sensitized surface of photographic film, or the image in that area; if numbered may act as a **locator**; 2. an individually numbered part of a **programmed learning text**

front matter—see **preliminary pages (prelims)**

granularity¹—level of **specificity** for the locator target in **eBooks**. The more granular the index, the smaller and more specific the sections of information it leads to. An index linking to sentence level is more granular than one linking to paragraphs.

grey literature—documents that are difficult to identify and not easily acquired through the normal book trade, e.g. conference literature; often synonymous with **semi-published documents** available only to a restricted group of people, e.g. committee minutes

hard copy—copy in physical form on paper, rather than electronic form as a computer file

heading—a term in an **index** representing a concept or item referred to in a document

hierarchy/hierarchical structure—an arrangement of **headings**, with subordinate **subheadings** and sometimes **sub-subheadings** for more specific concepts

homographs—words spelt the same but having different meanings, e.g. 'lime': 1. powdered limestone used in gardening/agriculture; 2. green citrus fruit; 3. type of theatrical lighting

homophones—words spelt differently but sounding the same, e.g. 'meat'/'meet'/'mete'

house style—a set of rules for spelling, abbreviations, **bibliographic citations**, **index** layout, etc. to ensure consistency within and between **documents**. There is some variation in how rigidly **publishers** and **editors** apply these.

HTML (HyperText Markup Language)—the markup language (code) used to enable material to be displayed in internet browsers. A form of HTML is now used to encode and display text in **eBooks**.

HTML5—a version of **HTML** developed between 2012 and 2014, extending it to include video, audio and scalable graphics, without the need for add-on software such as Adobe Flash.

impression—all copies of an **edition** printed at any one time and from one run of the press; all copies are identical. A new impression may incorporate minor corrections and revisions.

indent/indentation (indentation)—beginning or positioning a line (or lines) a certain number of spaces in from the margin. **Run-on subheadings** in print books are often indented from the left by 1 or 2 **ems**.

index—a sequence of **entries** systematically arranged so that users are guided to the information they seek in a document, a set of **documents** or a list of **documents**. It can be a whole **document** in itself, e.g. *Index Islamicus*.

index locator search—an option, available only with electronic text, whereby linked or embedded index terms relating to a particular section of text can be made to display at that point in the text. This gives the user more potential terms to explore in the index.

indexable element—a significant concept or item in a **text** which includes worthwhile information for the index user; see *also* **passing mention**

ISBN (International Standard Book Number)—a unique number that internationally identifies an **edition** of a book, e.g. ISBN 978 1 84354

586 6 is that of Christopher Hitchens, *God is not great*. London: Atlantic Books, 2007. Note that hardback and softback editions of the same work have different ISBNs.

ISSN (International Standard Serial Number)—a unique number that identifies a serial/periodical/journal, e.g. ISSN 0019-4131 is that of *The Indexer*.

journal—see **periodical**

keyword—a ‘significant’ **term** in a **document**’s title or text, e.g. ‘camping’ and ‘Brittany’ are keywords for *Camping in Brittany*, and could become **index headings**. Journal/periodical articles are frequently assigned keywords by their **authors** or **editors** and these may be used in computer-generated keyword **indexes**.

leaf—a single piece of paper with one page on each side; see *also* **recto**; **verso**

letter-by-letter—a system of **alphabetization** in which spaces, hyphens, dashes and diagonal slashes are ignored; see *also* **word-by-word**

locator (location reference)—the number and/or letter of a **page**, column, paragraph, line, section, file, **document**, or class (category) where information on a given topic appears. Locators in **eBooks** are clickable hyperlinks which take the user to the relevant text point.

looseleaf publication—a **document** that is constantly being updated, as **pages (leaves)** containing out-of-date material are discarded and those containing new material are inserted; particularly common for legal and financial publications

main heading—principal word or phrase forming the first element of an index entry, which may be followed by **locator(s)**, **cross-reference(s)** and/or **subheading(s)**

manuscript (abbreviation: **MS**)—literally a handwritten **document**; in publishing, normally the **document** or **copy** (whether handwritten, typed or printed) submitted to a **publisher** for editing and publication; may also be referred to as ‘typescript’

mashup—aggregate of existing indexes to a selection of books or **documents**, potentially making all their contents searchable through one index. While publishers are keen to use these, there are problems related to inconsistent approaches across indexes which can lead to inferior retrieval.

metatopic—the main topic of a **text**

microform—a miniaturized copy on film sheets (microfiche) or strips (microfilm)

minimal/minimum numbering—see **elision**

modifier—a word or phrase added to a heading to make it more specific and narrower in meaning, e.g.

birds, British

trade unionism, history

'British' and 'history' are the modifiers here (see *also* **qualifier**). A modifier may be presented as a **subheading** when there is more than one for the same **heading**, e.g.

birds

British

Scandinavian

monograph—a printed **document** that is not a **serial**, complete in one or more parts, devoted to one (often fairly narrow and specialized) subject

MS—see **manuscript**

multimedia document—a **document** using various types of mixed media, such as **text**, sound, still and moving images (photos, film, video), normally through a computer system, e.g. a **CD-ROM** encyclopaedia incorporating sound and moving images. Many **eBooks** are multimedia documents.

non-book medium—a **document** in other than printed, paged form, such as an **audiovisual document**, e.g. sound recording, film, **CD-ROM**, web pages, e-book, etc.

non-preferred term—a term from which a **see cross-reference** leads to a preferred term, e.g.

Beaconsfield, Earl of *see* Disraeli, Benjamin

'Preferred' means that for the purposes of a particular **index**, one term is chosen, though both may be used in the **text**.

oeuvre—the total output of an **author** or other **originator**

originator—a person or organization primarily responsible for the content of a **document**, e.g. **author**, photographer, film-maker, artist, map-maker, diarist, letter-writer, composer, sculptor, etc.

packager—see **book packager**

page—a single side of a leaf of a paper **document**, usually identified by a number or, occasionally, by a letter; also an individually identified and retrievable section of a computer-held **database** or a website

page number—common term and **publishers'** usage for normal page numbering of a book. The indexing term is **locator** or location reference, which covers page numbers, hierarchical paragraphs (as in these modules) and other types of numbering and symbols. In **eBooks** page numbers change as the **text** is reflowed, so these cannot be used as locators. Sometimes the page numbers from the print edition are still given, with these acting as hyperlinks, taking the user to an inserted anchor in the text where the information can be found

page proof—proof printed in **page** form before final correction

page range—cover-all **locator** indicating that subject matter is spread continuously over several **pages**. Only the first and last page numbers are given, usually joined by an **en-rule**, e.g. 407–412; see *also* **elision**.

pagination—allocating the **page numbers** to a **document**

passing reference/mention—an item or concept mentioned incidentally in the text but lacking worthwhile information about the item or concept itself; see *also* **indexable element**

PDF (Portable Document Format)—fixed-layout **document** format independent of computer software, hardware and operating system

periodical—a **document** published at (usually regular) intervals, e.g. a newspaper or journal

pinpoint indexing—in ePublications, indexing to the specific word or sentence where the reference begins

plugin—a small program that adds additional functionality or capability to another (often third party) software application

preferred term—a **term** used to represent a concept used in preference to another synonymous or near-synonymous concept, e.g.

periodical *see* journal

where 'journal' is the preferred term; see *also* **non-preferred term**

preliminary pages (prelims)—the parts of a book preceding the main **text**, often numbered in small roman type, e.g. title page, contents list, introduction, preface, foreword, acknowledgements; often referred to as front matter

printer—a person or organization printing **documents**, often employed by a **publisher**

programmed learning text—**document** presenting information with questions, the answers to which determine the user's progress through the **document** and the order in which the sections of **text** are read

proof—a printed copy of the **text** of a **document** before final corrections to **text**, layout and pagination

proper name—a **term** referring to and identifying an individual person, organization or place, e.g. Agamemnon, Marks and Spencer, Shropshire. The words tend to begin with upper-case (capital) letters.

publisher—a person or organization first making a **document** publicly available

qualifier—a word or phrase that is part of the **heading**, frequently in parentheses (round brackets); used to add more explanation or to distinguish between homographs or identical names, thereby removing ambiguity, e.g.

mouse (animal)

mouse (computer device)

Taylor, Elizabeth (actress)

Taylor, Elizabeth (writer)

Paris, France

Paris, Ontario

see also **modifier**

recto—a right-hand **page**

reference—see **bibliographic citation**

reflowable text—text in an **eBook** file that is reformatted always to fit on the screen regardless of selected text size. Scrolling to the left or right is not necessary. Pages are not therefore fixed, and this is the reason a standard paginated index cannot be used when a print book is converted to an eBook. **cf. PDF**

related term—a **term** that is closely related to another and to which a **cross-reference** is made

repurposing—re-using the **text** of a **document** e.g. reformatting a document from one form to another such as hardback to **eBook**, or selecting parts of the document to create a new publication

revised edition—an **edition** of a **document** in which the content has been significantly changed from the previous **edition**

RTF (Rich Text Format)—a **document** file format for cross-platform **document** interchange, commonly used to deliver index copy to clients

run-on subheading—a **subheading** starting on the same line after a **main heading** or a previous **subheading**. Depending on the column width, if there are several **subheadings**, **turnover lines** will be required; these will probably be indented, e.g.

bread: breadsticks 5, 9; brown loaves 17, 23, 27;
unleavened 3, 35, 37; white loaves 12, 14, 15

search term—term used by an enquirer using an **index** to seek information on a particular topic

see also cross-reference—an instruction directing the index user from a **heading/subheading** with **locator(s)** to one offering related information, e.g.

jazz 32–40; *see also* blues

see cross-reference—an instruction directing the index user from a **heading/subheading** without **locators** to one that may be a **synonym** or near-synonym or a **related term**, when the same concept may appear in several different ways in a **text**, e.g. in a book on music:

Afro-American music *see* jazz
Dixieland music *see* jazz
New Orleans music *see* jazz

self-indexing document—a **document** which is arranged in alphabetical order, such as a dictionary, encyclopaedia or directory (but may still need an analytical index)

semi-automated indexing—analysis of text by software programs that use occurrence based systems with inbuilt weighting for significance

self-published document—increasingly, with the advent of **eBooks** some **authors** are publishing directly to the web themselves e.g. via Kindle

semi-published document—see **grey literature**

serial—a **document** issued indefinitely in successive parts, each part bearing the same title but identified separately by number or date, e.g. a newspaper, a **periodical** (magazine, journal), a yearbook, an annual report

series—a set of **documents**, each having its own title but also identified by a collective title for the set, e.g. the Arden Shakespeare edition of separate Shakespeare plays; television documentaries under a general title, such as *Horizon* or *Equinox*; Society of Indexers' Occasional Papers on Indexing, which cover different aspects of the subject

set-out subheading—a **subheading** presented on a new line and indented one or more spaces (**ems**); not continued on the same line as the main heading or previous **subheading**

specificity—the precision with which a concept or topic is identified in an **index**; influenced by the **exhaustivity** of indexing

standalone indexing—conventional indexing where the index file is supplied by the indexer as a separate file from the text file. The term is used to make a distinction from **embedded indexing**. A different understanding is used in the EPUB3 charter where it refers to a publication that consists only of one or more indexes to other **EPUBs** or external targets

string—a greater number of **locators** after a heading than the index user can easily handle. Six **locators** on average is regarded as acceptable; if there are more, it is advisable to create subheadings from them to give easier access to information. Long strings may occasionally be unavoidable, especially after a subheading when only one level of **subheading** has been specified by the client, but they are to be used only if all else fails. Strings may need to be shorter in **eBooks** for clarity (i.e. where each **locator** is indicated by the same clickable symbol), necessitating more subheadings, and in some cases there may be a separate subheading for each **reference** with the subheading itself acting as the clickable link

subentry—a record in an **index** consisting of a **subheading** with **locator(s)** or with a **see cross-reference**

subheading—a **heading** subordinated to a **main heading**, representing a subdivision or particular aspect of a topic; see *also* **run-on subheading**; **set-out subheading**

sub-subheading—a **heading** subordinated to a **subheading**, representing a further subdivision or more specific aspect of a topic

superior term—a **term** higher up in a hierarchy of **terms**, especially in a **thesaurus**

synonym—a **term** having the same meaning as another **term**, e.g. food/nutrient

term—a word, phrase, symbol or set of symbols (as in music, mathematics or science) used to represent a concept

text—1. the information content of a **document**; 2. information recorded in letters or numbers (written, typed, printed) rather than images and sounds

thesaurus—a controlled and structured list of **terms** for use when indexing **documents** and retrieving information, e.g. *MeSH* (Medical Subject Headings), which is used as an authority for headings for *Index Medicus*

title—the unique name identifying an individual **document**, e.g. *Gone with the wind*; *The mikado* or *the town of Titipu*. The fullest version is found on the title **page** rather than the cover of a book. When indexing, always follow the usage of the **text**, especially with regard to capitalization.

turnover line (turnover)—the second (and subsequent lines) of a paragraph or index **entry**. In a printed **index**, turnovers must be indented more deeply than any other **indentation** used for **subheadings**, etc. e.g.

pies

 blueberry pies 34

 Cindy Lou's Special Thanksgiving

 Huckleberry Pies 35

typesetter—a person or organization preparing **text** for printing

verso—a left-hand page

word-by-word—a system of **alphabetization** in which spaces, hyphens, dashes and diagonal slashes between words are given filing values (as in this list of definitions); see *also* **letter-by-letter**

workflow—a term used in the publishing industry to describe the sequence of events that takes place in the publishing production process

XHTML—this is essentially a stricter form of **HTML**, meeting **XML** standards of well-formedness. XHTML is supported by a W3C Recommendation and is the basis, for example, of the EPUB standard for **eBooks**

XML (Extensible Markup Language)—a markup language designed to transport and store information, primarily **documents** and publications, in which the markup code can be designed individually for the project (unlike **HTML**, a form of XML for internet pages, where the coding is standardized). Some publishers now code the **text** in XML and use this to output to an **eBook** format

Index functions and characteristics

Introduction

This section introduces you to:

- what indexes are for and why we need them
- the main features of indexes – entries, headings, subheadings, cross-references and locators
- the conventions of index form and layout
- how index entries are compiled.

These topics are covered in more detail in subsequent sections of this and later modules.

A2.1 Function

An **index** provides a way of finding information in a **document**, a set of documents or a list of documents. It is an aid to finding references to major treatments of topics, and to lesser items that may be of interest in another context. It brings together scattered references to different aspects of the same subject and links **related terms** and **synonyms**.

Without an index, the user must either work through the whole document (or set or list) from beginning to end; or dip into it at random, looking for the wanted information; or try to guess (from contents lists, chapter headings, captions, words in **titles** or other indicators) where it might appear.

A contents list at the front of the **text** is valuable as a quick guide, and is always worth studying, but it lists only the principal divisions of the text in the order in which they appear, and gives only a broad indication of the textual material. An index rearranges the information in the whole document, giving access to all of it.

The majority of indexes use words as **headings**, although some may use only numbers, as with telephone directories that are arranged by

Look in **Section A1:** 'Definitions' to find the meanings of words highlighted in bold type at their first mention.

See A1.1 for reasons why indexes are still important in **eBooks** which also provide text searching facilities.

number (beginning at 0 – nought/zero) so that users can find the place name if they know the telephone number. Other indexes use symbols as headings (e.g. mathematical and scientific symbols). This module deals with languages (alphabetical letters) only.

A2.2 Necessary characteristics

Every index contains **terms** arranged in a sequence of **entries**. An entry may consist of

- a heading (i.e. a **main heading** with or without **subheading(s)**), at least one **locator** and sometimes a **see also cross-reference** to at least one related term; or
- a main heading or a subheading with a **see cross-reference** to a **preferred term**.

A2.2.1 Main headings and subheadings

Indexes are commonly set out in an alphabetical sequence:

animals
musical instruments
water plants

These are main headings and are aligned at the left-hand side of the **page** or column. There may also be subheadings **indented** (usually by a **1-em** space) under the main headings. This convention is followed throughout these modules.

animals
 breeding grounds
 feeding habits
musical instruments
 history of development
 manufacture
water plants
 garden ponds
 rivers

Subheadings may be **set out** in a list, as above or run on on the same line and separated by semicolons:

animals: breeding grounds 26–38; feeding habits 39–57

Set-out subheadings are used in eBooks.

Sub-subheadings are similarly indented and set out under subheadings:

Africa
 animals 26–57
 breeding grounds 26–38
 feeding habits 39–57

If it is necessary to make them run on, they can be enclosed in parentheses (curved brackets):

Africa: animals 26–57 (breeding grounds 26–38; feeding habits 39–57)

A2.2.2 *See also* cross-references

Sometimes a *see also* cross-reference to a related term is useful or necessary:

musical instruments *see also* dance bands; orchestras

While indexing a general work on musical entertainment, the indexer has helped the user by connecting subjects from different parts of the text. Here the different ways of combining musical instruments are being pointed out by the *see also* reference; the words or phrases that follow it are known as related terms.¹

A2.2.3 *See* cross-references

Sometimes a heading is rejected in favour of its synonym or usage in another language:

duckweed *see* *Lemna minuta*

Both terms may be used in the text of a book on water gardens, say, but the indexer gradually discovers that the Latin names of plants predominate and so to be consistent prefers the Latin term in the index. ‘Duckweed’ must be mentioned, as it does occur in the text and some users are likely to look for it, but it is the **non-preferred term**; the *see* cross-reference directs the user to the Latin, which is the preferred term in this particular index.

Note that a *see* cross-reference is not needed if the preferred term would be adjacent in the index:

personal development plans 4

PGCE (Postgraduate Certificate in Education) 192

primary school teaching 192

(no entry needed for Postgraduate Certificate in Education as this would be adjacent)

The Latin names of species are set in italics.

A2.2.3.1 *Double entry*

Double entry is often used instead of *see* cross-references if there are only one or two locators, for example:

training, vocational 33, 42–48

university courses 25, 49

validation 64–68

vocational training 33, 42–48

or if the entry does not occupy more space than a cross-reference will. It is also helpful as it prevents the index user from having to turn over

¹ For more information on the layout of subheadings see **Module D Section D2.3.1**.

many pages of an index to reach a term that is alphabetically a long way from the original heading. For example, instead of:

engineering, mechanical *see* mechanical engineering

use double entry:

engineering, mechanical 272–320

mechanical engineering 272–320

A lot depends on the type of index, however, and if your preferred term has many subheadings it would be wiser to refer to it by a *see* cross-reference, as in:

training, vocational *see* vocational training

university courses 25, 49

validation 64–68

vocational training

nursing 33

primary school 42–45

secondary school 45–48

theological 93–95

so that users may choose the subheadings they require.

In eBooks the use of double entry of terms is not constrained by space considerations.

A2.2.4 Locators

Locators (location references) may consist of **page numbers** (or letters) or the numbers (or letters) of sections, paragraphs, columns, lines, **frames**, documents, files, shelves, classes, or any other suitable identifiers. All numbers move from low to high; prefaces and other **preliminary pages** (prelims) may use Roman numerals, but the main text invariably has Arabic numerals. In the index it is usual for Roman locators to precede the Arabic, for example:

vocational training iii, xxvii, xlvi, cxxxii, 3, 27, 46, 132

Locators referring to illustrations, maps or diagrams in books may be highlighted by using a different typeface (often italic) and may be placed either in the proper sequence or at the end of it:

vocational training 3, 27, 39, 46, *115*, 132

or

vocational training 3, 27, 46, 132, 39, *115*

There are various ways of indicating illustrations in an index, and it is advisable to find out the client's preference.

When numbers and letters are used together as locators (as might be done with the paragraphs of these modules), the numbers come after the letter in the locator sequence:

vocational training A2.3.1, B2.1.3, C1.3.2

Here the letters represent larger divisions of the text than do the numbers. If larger divisions are numbered and smaller ones use letters, then the locators would naturally be in reverse:

vocational training 3.a.b, 4.b.c, 5.c.d

When a concept is dealt with over several pages of a document without interruption, it is usual to give in the index only the **page range** (i.e. first and last page numbers), joining the numbers with an **en-rule**. Such locators may be given in two ways: in full or in **elided** form (also known as **minimal numbering**). The indexing standard BS ISO 999² recommends that the full form of numbering should always be used for maximum clarity:³

vocational training 36–39, 115–118, 123–124, 132–147, 150–152

However, professional indexers may be asked by clients to use the elided or minimal form to save space:

vocational training 36–9, 115–18, 123–4, 132–47, 150–2

If a concept appears on consecutive pages but the discussion is constantly interrupted by other textual material, its locators must be shown in the index quite separately and not combined:

vocational training 36, 37, 38, 115, 116

Abbreviations indicating items in a **bibliography** or appendix are often used as locators in indexes to **periodical**/journal articles, for example:

Matheson, William 49, 438Bib.

Springfield, Mary 38, 206Ap

Here 'Bib.' refers to the bibliography and 'Ap' refers to the appendix.

Indexes to older publications may contain the following Latin terms next to locators:

14 *et seq.* 36ff. 126–143 *passim*

The first two mean 'and on pages following' but neither offers an end point to the mention of the subject; '*passim*' means 'throughout, dispersedly, here and there' and is used to cover sporadic mentions of the same subject. These terms are now felt to be too vague as they do not provide exact locators, though indexers have been known to have long discussions about the possible virtues of *passim*.

Note that in elided numbering, the teens should never be too minimal: i.e. 115–18, not 115–8, since 8 on its own here is less easily read.

2 British Standards Institution, 1996. *Information and documentation: Guidelines for the content, organization and presentation of indexes*. London: British Standards Institution (BSO ISO 999: 1996)

3 See the Resource Centre for details of the British Standards referred to throughout all modules.

A2.2.4.1 Locators in tagged indexes

In a tagged index for an unpaginated document the reference that the index term points to is marked in the text by a code. This code then becomes the locator in the separate standalone index. The tags are transformed into page numbers/ranges at a later stage of book production. This method of numbering allows the **publisher** to 1) alter the **pagination** after the index has been completed and 2) in some cases to 're-purpose' the index e.g. from hardback to paperback.

See examples of tagged and embedded indexes in Figure A1.

A2.2.4.2 Locators in embedded indexes

In an embedded index, the index entry is actually placed (embedded) in the text itself at the appropriate point. In MS Word, for example, the entry is embedded in the **XML** source code of the document by the software. Pagination does not need to be set at this point. The software will generate the index from the embedded entries and create the correct locator for the final pagination during the process.

There is more on the differences between tagged and embedded indexing on the SI PTG website <http://www.ptg-indexers.org.uk/resources/notes.htm>

A2.2.4.3 Locators in eBooks

Page numbers do not work as locators in eBooks. The eBook is essentially one very long page, or series of very long pages (like a very long webpage). The 'screen' page the reader views changes as **font** sizes are changed and text is **reflowed**. A 'page' is therefore not a stable reference point.

See the presentation on 'Understanding eBooks' in the Resource Centre of the SI indexer-training website

Locators in eBooks should be active links. The reference point is created by introducing anchors into the text. This can be done at section, paragraph, sentence or word level. This is referred to as the '**granularity**' of the level of indexing. Sentence or word level, known as **pinpoint indexing**, is the most precise and the hyperlinked index entry takes the reader to exactly where the reference is displayed on the screen page.

There are various ways of doing this but most indexers are unlikely to get involved directly in this process; some publishers use XML-based **workflows** in the production process which allows use of embedded indexes (indexes generated from entries embedded in the text) or tagged indexes (indexes supplied separately with entries linked to identifiers in the text) to create active linked indexes.

Format of locators

Each locator may be represented by a sequential number or a symbol which is actually a hyperlink to the text: clicking the link takes the user to the text. In some cases, each locator is given its own subheading and the subheading itself is the clickable link.

Where an eBook is produced alongside a print **edition**, the locator may utilize the page number from the print edition as the symbol. It has no meaning in terms of actual location and just acts as a link to the anchor.

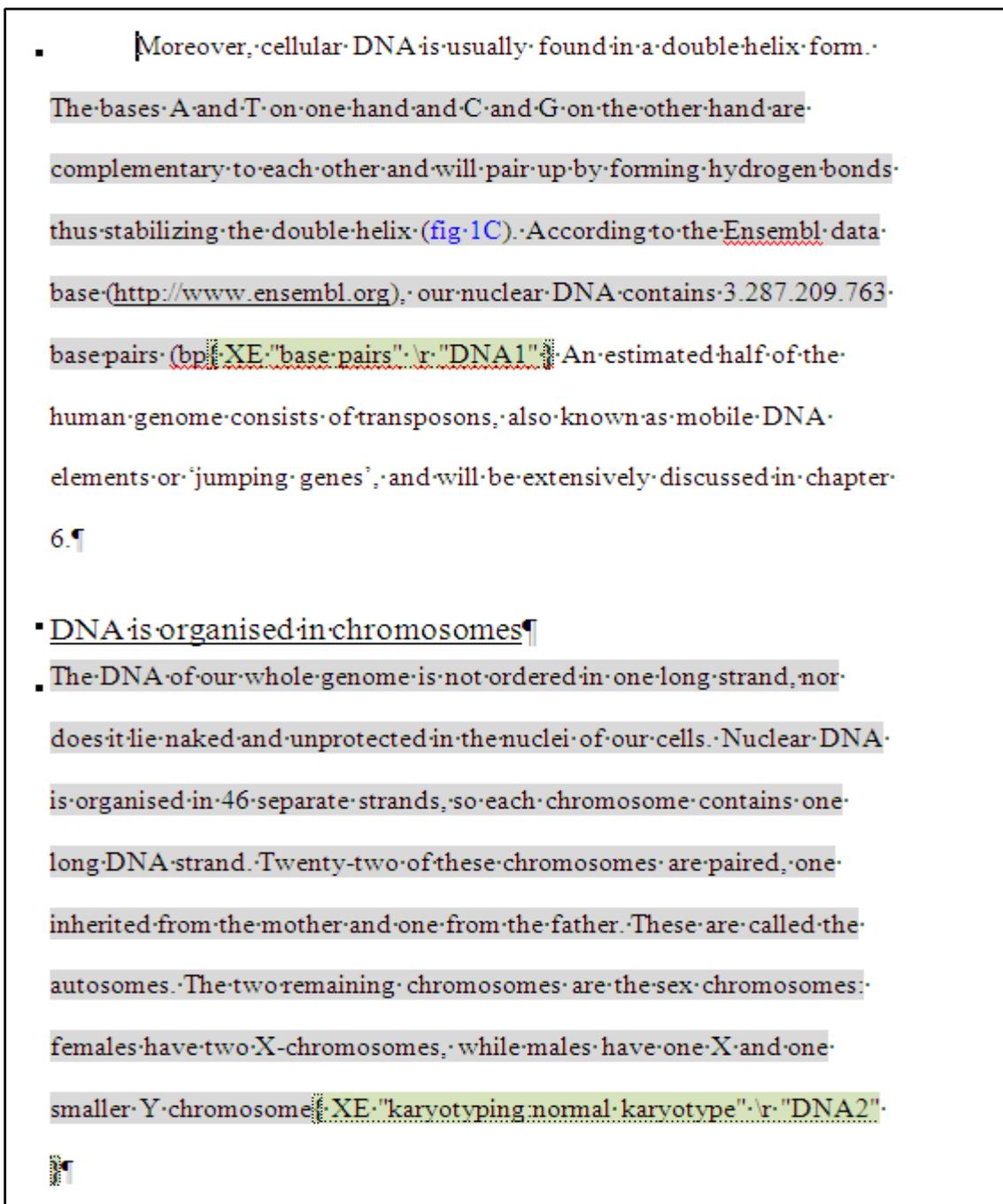
Para	<p>1 Moreover, cellular DNA is usually found in a double helix form.</p> <p>2 17.20b The bases A and T on one hand and C and G on the other hand are</p> <p>3 complementary to each other and will pair up by forming hydrogen bonds</p> <p>4 thus stabilizing the double helix (fig 1C). According to the Ensembl data</p> <p>5 base (http://www.ensembl.org), our nuclear DNA contains 3.287.209.763</p> <p>6 base pairs (bp) f17.20. An estimated half of the human genome consists of</p> <p>7 transposons, also known as mobile DNA elements or 'jumping genes',</p> <p>8 and will be extensively discussed in chapter 6.</p>
B	<p>9 <u>DNA is organised in chromosomes</u></p>
Para-no-indent	<p>10 18.100b The DNA of our whole genome is not ordered in one long strand,</p> <p>11 nor does it lie naked and unprotected in the nuclei of our cells. Nuclear</p> <p>12 DNA is organised in 46 separate strands, so each chromosome contains</p> <p>13 one long DNA strand. Twenty-two of these chromosomes are paired, one</p> <p>14 inherited from the mother and one from the father. These are called the</p> <p>15 autosomes. The two remaining chromosomes are the sex chromosomes:</p> <p>16 females have two X-chromosomes, while males have one X and one</p> <p>17 smaller Y chromosome. f18.100</p>

Index entries tagged using manually inserted identifiers.

Tags are **highlighted** for the publisher's benefit and the index is compiled and delivered separately using the tagged references as locators. Each range of text has a beginning (b) and a finishing (f) tag.

Figure A1 Screenshots showing tagged index and embedded index versions for two entries. 'base pairs' and 'karyotype, normal'

Screenshot one: tagged index version



Index entries embedded in text using MS Word Index function.

The embedded entry is normally hidden but is shown here highlighted. The range of text is defined by a Word *bookmark* [DNA1; DNA2] and the entry connects to this using the code \r

Figure A1 continued Screenshots showing tagged index and embedded index versions for two entries. 'base pairs' and 'karyotype, normal'

Screenshot two: embedded index version

A2.3 Common practice: conventional usage

The annotated examples set out in Figure A2 are taken from different indexes. They are not in alphabetical order but they indicate some of the forms and layouts in which index entries can appear. They represent the practice recommended in this course and in the indexing standard BS ISO 999.⁴ You will learn more about these various types of entries, headings and locators as you progress through the course.

Capital letters and small letters

It is usual to begin headings that are ordinary words with a small or lower-case letter. **Proper names**, titles of works, or words and phrases that are special in some way normally begin with a capital or upper-case letter, for example:

Villa-Lobos, Heitor
 violas
 violins
 violoncellos
Vissi d'arte (Puccini: aria)
 Vivaldi, Antonio

This is now regarded as the standard style. However, some British **publishers** and many US publishers prefer to have all index headings beginning with capital letters; the disadvantage of this is that it can lead to ambiguity with **homographs** – China (the country) and China (pottery).

Punctuation

Punctuation is only used where necessary; the 'Tate Gallery' entry in Figure A2 has none; but the entry for 'fats', with its run-on subentries, is made clear to the index user by including a colon and a semicolon. Throughout this training course and BS ISO 999 there is a double space with no comma between a heading and its first locator/cross-reference, although the **house style** of some publishers may require a comma in this position.

Indentation

Indentation by a 1-em space is normal for lines that begin further in from the main heading, as with subheadings and cross-references, with a further 1-em space for sub-subheadings. Occasionally clients require a larger indent, but whatever the level of indentation used it must be

BS ISO 999: 1996, clause 7.2.2.3, recommends using lower case to begin all ordinary words.

Upper case and lower case originally referred to the long wooden cases of pigeonholes in a printer's shop which held moveable lead type. Capital letters were stored in the upper case – in alphabetical order – and small letters in the lower case: hence the usage upper-case/lower-case letters, often abbreviated to uc/lc. Woodcuts and photographs of these cases may be seen in histories of printing.

4 British Standards Institution, 2005. *Marks for copy-preparation and proof correction*. London: British Standards Institution (BS ISO 999).

French Revolution 34, 45–49	<i>Main heading; locators in full form</i>
lenses	<i>Set-out and indented subentries; locators in full form; locator for illustration in italic</i>
aerial photography 123–124	
architectural photography 114	
natural history photography 110	
portraiture 125, 126	
Nigeria	<i>Set-out and indented subentries and sub-subentries</i>
economy 78–81	
languages 104–105	
population	
growth 23	
mobility 75	
Single European Market 3.1.4	<i>Paragraph number as locator</i>
automobiles <i>see</i> cars	<i>Cross-reference directing the user from non-preferred term to preferred term; a ‘see’ reference does not normally include locators</i>
nuclear energy 56-57, 89 , 145	<i>Locator for major textual reference in bold type; locators in full form; set-out subheading and cross-reference to related term; deepest level of indentation used for turnover lines</i>
bibliography 275–278	
<i>see also</i> electricity supply industry	
flour 36, 57; <i>see also</i> wheat	<i>Run-on cross-reference</i>
heraldry 929.6	<i>Classification number as locator</i>
Ashton-Warner, Sylvia 14a	<i>Hyphenated family surname in preferred order as part of heading; page number and column letter as locator</i>
Smith, Nancy Banks- <i>see</i> Banks-Smith, Nancy	<i>Second part of hyphenated family surname as first part of heading, but as a non-preferred term</i>
Henry III, King of England 153–8	<i>Name of royal as keyword, not preceded by title (i.e. not King Henry)</i>
fats: analysis 130–131; types 156	<i>Heading with colon; run-on subentries separated from each other by a semicolon</i>

Figure A2 Some examples of the many different types of index entries and index layouts (continued on p. 25)

consistent throughout the index. A **turnover line** (where text overruns a specified column width and continues on the next line) should be indented further than the deepest indentation used for headings and subheadings.

In **eBooks**, indexes are usually presented in a single column as there is no space restraint.

Bold type

Bold type may be used to show which locator in a sequence refers to the most important material about a concept. It can also be used for headings or important subheadings in indexes to large publications, e.g. encyclopaedias, law indexes.

cloud: how sweet to be a c. 64:9	<i>Keyword in quotations index abbreviated in quotation; page number and item number as locator</i>
<i>Mary Rose</i> (Tudor ship) 59	<i>Name of ship in italic; qualifier in parentheses as a clarifying explanation of what could be the title of a literary or artistic work</i>
Constable, John: <i>The hay wain</i> 16–21, 19	<i>Title of painting in italic; locator for illustration in italic, of necessity out of sequence</i>
<i>Whitaker's Almanack</i>	<i>Title of publication in italic</i>
Tate Gallery T6	<i>Filing system number as locator</i>
education (see also learning resources) policy 86/94, 87/346; see also local education authorities sociology of 88/92, 88/1014	<i>Cross-references to related terms from heading and set-out subheading produce turnover lines more deeply indented than the subheadings; years and document numbers as locators</i>
education Adult education survey EdN Aug 90 4–5	<i>Title of journal article with abbreviated bibliographic reference as locator</i>
office layout The virtual office: putting management ahead of facilities (Stocks). <i>Facilities</i> 16 (1/2) Jan/Feb 1998 29–33	<i>Entry in an index to periodicals, consisting of a heading and a subheading causing several turnover lines and comprising subject, article title, author in parentheses, journal title in italic, volume number in bold, part number in parentheses, date and inclusive locators</i>
Value Added Tax see VAT	<i>Cross-reference directing reader from full phrase to its acronym as preferred term</i>
VAT (Value Added Tax) 63–65	<i>Acronym followed by full form in parentheses and by locators</i>
Gordon, C. <i>Teaching the young to use indexes.</i> 13 (3) 1983 181–182	<i>Entry in a volume index to a periodical [The Indexer], consisting of author, article title in italic, volume number in bold, part number in parentheses, date and inclusive locators; turnover lines indented</i>

Figure A2 continued

Because **typesetters** have different coding practices, when sending indexes in electronic format only, it is wise to check that bold and italics will be transmitted correctly. Discuss with your client any codes required before starting the index. Before the advent of sophisticated word processors able to reproduce different typefaces, underlining was used to draw the typesetter's attention to text to be set in a different **font**: i.e. straight underlining for italics, wavy line for bold. Although it is now rarely necessary to mark **hard copy** in this way, you will need to be familiar with the standard proofreading marks, which use similar conventions. BS 5261C: 2005 sets out the standard marks used in both **copy**-preparation and **proof** correction.⁵

5 British Standards Institution, 2005. *Marks for copy-preparation and proof correction*. London: British Standards Institution (BS 5261C: 2005).

Italic type

Italic type is generally used for titles of publications, names of ships, paintings, foreign words, locators of illustrations and for *see* and *see also* cross-references. In biology, italics are used for Latin binomials, but not for names of viruses.

House style

Figure A2 illustrated a variety of index entries, but other forms of entry can also be found in indexes: commercial publishers frequently have their own individual styles – house styles – which they use consistently in all their publications and which an indexer may be expected to follow. Activity A1 highlights some of the features you should look at when exploring printed indexes.

Activity A1

Look at as many indexes as possible, in your local library or your own collection, to see how different forms of entry and layout are used. Pay particular attention to the following, all of which may vary according to house style.

- Spacing. One space or two after a heading and before locators? How many spaces used for indenting subheadings? Are there linespaces between alphabetical sections?
- Punctuation. Comma or no comma between heading and first locator?
- Capitalization. Are the first letters of main headings capitalized or not?
- Highlighting. Is bold type used for important headings or locators?
- Layout. Are subheadings set-out or run-on?
- Cross-references. Where are *see also* cross-references placed – before or after subheadings?
- Locators. Are they in full or elided form – 317–319 or 317–19?

A2.4 Indexable elements: what to put in the index

The compilation of indexes is an intellectual process that demands, above all, exactness – of copying, spelling, spacing and punctuation. A text/document must therefore be thoroughly understood and the index compiled from it must be comprehensible to the index user and be as exact as possible.

The document must first be examined for **indexable elements** – significant concepts and items in the text for which worthwhile information is given. These must then be transformed into suitable index headings: some headings can be taken straight from the text; others, such as some types of subheading and cross-references, must be developed in the mind of the indexer. For example, for an informative work on the development of domestic architecture, the indexer may

In identifying what to index and developing headings and subheadings, you must constantly ask yourself: 'What will the user be looking for?'

have to consider types of stone and where they come from, types of earth for brickmaking, types of timber and metal for structural work, and furnishings such as chimneys, stairs, doors and windows – and how all these change through the ages. Many words or concepts in the text will immediately be seen to be appropriate headings for the index, but condensing the descriptions in the text to form concise subheadings may require some thought. Figure A3 gives some examples of how descriptions can be condensed into very precise terminology.

<p>chimneys (see also andirons; fire-grates; fireplaces; inglenooks; mantels; overmantels) draught control and 124 flues branched 350 single vertical 348 at inner wall 408 at outer wall 407 as smoke outlets see chimney pots</p>	<p><i>Main heading with set-out subentries and sub-subentries in alphabetical order. Note that linking terms and prepositions are ignored in alphabetizing. In this instance the cross-reference directing the user from the main heading to related terms is placed immediately after the main heading, the terms being separated by semicolons, and parentheses are used to avoid confusion with subheadings.</i></p> <p><i>Cross-reference directing the user from subheading to preferred term; note that turnover lines are indented further than the deepest subheading.</i></p>
---	--

Figure A3 Examples of subheadings that might be devised for an architectural index

It is helpful to identify the **metatopic**, or main topic of a text (in some texts there may be more than one). An index entry for the metatopic is not usually needed, but in certain circumstances it can be useful.⁶

Mere word-spotting – underlining important words in the text and listing them alphabetically – shows lack of thought; it reveals that the text has not been thoroughly examined and may lead to howlers. H. B. Wheatley relates the possibly apocryphal story of an index entry ‘Best,

Mr Justice, his great mind’, referring to a passage of text reading: ‘Mr Justice Best had a great mind to commit the man for trial’. In this case, however, the entry may have been intended as a slight against the judge rather than the result of carelessness. Many of H. B. Wheatley’s examples of bad indexing practice and poor term selection are as apposite today as when they were first written, over 100 years ago. For full details of these publications, see Resource Centre.

Implied concepts

Indexers need to be aware that it is important to analyse the text to identify the underlying unnamed concepts, as well as those named in the text. Index entries and cross-references need to be provided for all topics, whether named or implied. This can make the difference between a good index and a poor one.

To help the user, some of the subheadings in Figure A3 should also appear as main headings, for example, ‘draught control’ and ‘flues’.

⁶ See **Module B Section B1.3.1.1** for more information on metatopics.

Consider the following reference:

‘Test matches were played here in the 1950s’

This clearly refers to cricket, but that is not actually stated. A user looking for references to cricket will not be led here unless the indexer has identified the underlying topic and included an index entry.

Another example, from the *Sussex Archaeological Collections*, vol 151 (2013), p. 117:

‘Also in 1648, Horsham was the centre of a royalist uprising suppressed by Sir Michael Livesay’s parliamentary regiment of horse. There was fighting close to the school, and the quartering of Livesay’s troops in the town during 1649, characterised by “disorders and plunderings without distinction of friend or enemy”, caused great distress.’

This refers to the English Civil War, but unless the indexer has sufficient knowledge and skill to recognize this and provide an appropriate entry, the reader looking for information on the Civil War will not find it.

Here is an example of a scientific subject:

‘Blood pressure should be adequately controlled with maximally tolerated alpha-blockade. Different regimens are in use to achieve this objective. Our practice is to start patients on phenoxybenzamine (an irreversible alpha-blocker) at a dose of 10mg twice a day.’

This refers to the control of hypertension with antihypertensive drugs, but neither of these terms appear in the passage.

This is one of the strengths of human analytical indexing over automated search procedures.

Indexable elements may be proper names, such as those of individual people, organizations and places, or you may be dealing with categories such as activities, manufactured products, abstract ideas, qualities of a person’s character, animal life, natural and applied processes, techniques, skills and opinions. Some indexes may be very unspecialized, as for simple advisory books with a wide readership, perhaps on childcare, cookery, or common ailments of old age. Then there is a large area of middle-ground works, including biographies/autobiographies, histories, and scientific and medical books for the layperson. At the other end of the spectrum are works that demand specialized knowledge in addition to indexing knowledge; to tackle post-structuralism in literary theory, parasitology, or tropical diseases, an appropriate university degree would be an advantage, if not essential.

Whatever your level of knowledge, and however general or specialized the text, the basic method of indexing and the construction of the index remain the same.

Section A4.3 below describes some of the qualities required to make a person a good indexer. Some experienced indexers have written about their methods of working and their own approach to the indexing process. See Resource Centre: How I Index.

A2.4.1 Thesauri and classification schemes

A **back-of-the-book index** normally uses terms found in the text of the document, supplemented by terms that the indexer expects potential users to look for. Additional relevant terms can be discovered by consulting an appropriate **thesaurus** or a subject heading list; one may be necessary, for instance, to help with an index to a work on the historical and technical aspects of photography.⁷

Other specialist indexing techniques are practised in environments such as libraries and resource centres where the stock is arranged according to a **classification scheme**, and in subject areas with a large periodical and report literature, for instance, engineering, chemistry, law, electronics, social sciences and medicine.

A2.5 Direct entry and classified indexing

Most indexes have three basic components: main headings, subheadings (which may include sub-subheadings) and cross-references. There are two ways of handling these and an indexer may be asked to use either method.

Direct entry

An index in direct form is usual for most works for the general market and for many specialized works. The majority of terms appear as main headings, sometimes with subheadings and cross-references. All the examples in these modules are given as if for direct entry indexing. The index user can look up anything from the text and find it either at once or one step away by means of a cross-reference.

Classified indexes

Classified indexes were traditionally used for works on law (and sometimes in the sciences), although there are many indexes in these subjects using direct entry. In financial law, say, there can be a process of continuous indexing where the main headings remain the same for many years, but the Chancellor's annual Budgets continually alter the finer details, so that the great list of subheadings for each main heading may not only need adjusting but also adding to. Thus, subheadings predominate to the extent of many columns and main headings are often printed in bold type to catch the eye. In a medical index, the main headings might consist of the names of diseases and the subheadings of

⁷ The use and construction of thesauri are covered in **Module B Section B2.9.1**: 'Using thesauri' and **Module D Section D3.3**: 'Thesaurus construction'.

drugs used for treatment. But research continually causes the chemical combinations to alter, so here again, long lists of subheadings keep changing and/or getting longer. Such indexes are not expected to be consulted in a hurry; it may take some time for the user to look down the lists to find the required heading. These indexes tend to belong to reference works that are to be found in institutions and specialist organizations, such as law firms or medical libraries.⁸

For the purpose of indexing in general, most people are expected to create indexes in the direct entry form. A classified arrangement may lead to important information being hidden under main headings where a reader is unlikely to look, for instance:

mammals
 antelopes 63–4
 horses 76
 stoats 35

A good example of dense legal indexing using classified entries is 1999 *Halsbury's Statutes*, 4th ed. London: Butterworth.

A2.6 The position of the index

An index is normally produced as an integral part of the document to which it belongs, as with a typical back-of-the-book index. An online index is an integral part of an electronic document, such as a website index. An index can also be bound separately, as with the indexes to *Encyclopaedia Britannica*. Another type of separate index is a journal index, usually produced annually and covering the whole year's articles; these are often bound with the final journal issue of the volume. Many journals have **cumulative indexes** covering periods of, say, five or ten years, or sometimes much longer. A separate index can also appear in a different physical format from the 'parent' document: a good example is the *Index of Christian art*, compiled by Erwin Panofsky and others, to photographs of religious works from all over the world, available online at <http://ica.princeton.edu>.⁹

Indexes in eBooks may not necessarily be at the back of the book. Ideally an easy access link should be provided from any point in the text. In some more complex texts or book apps there is a sophisticated search tool which uses an underlying index construction or thesaurus to provide focused search results.

See the article on Bartlett's familiar quotations app in *The Indexer* 31 No 3 pp122-3.

A2.7 Single and multiple indexes

Although recommending that there should be a single index to the contents of a single document, rather than a series of indexes, BS ISO 999 does recognize that there may sometimes be a need for more than

BS ISO 999: 1996, clause 7.1.4 recommends that there should be a single index.

⁸ See **Module B Section B3.1.4** for more information on classified indexing.

⁹ There is more about this in **Section A7.1.5** below. See also **Module D Section D4.2**: 'Cumulative indexing' for more about the particular problems of indexing long runs of journals.

one index.¹⁰ Users may have a particular interest in only certain parts of a document, or certain types of information. Examples of separate sequences provided for ease of reference include: advertisements and book reviews in periodicals; **authors** separated from titles and subjects in a detailed history of literature (or in a collection of abstracts of scientific papers); cited legal cases in a law text; and patent numbers kept separate from word indexes attached to the same work.

A2.8 Summary: types of index entry and arrangement

After studying this section you should now be gaining some familiarity with indexing terminology and understand that indexes normally contain alphabetically arranged entries consisting of headings or subheadings in the form of:

- **personal names** (usually with family name first) – Branson, Sir Richard; Laine, Cleo
- **corporate (organizational) names** – Commissioners of Northern Lighthouses; John Lewis Partnership
- **names of places and topographical features** – Kalahari Desert; Sheffield
- **titles of documents** (books, songs, works of art) – *Business studies for GCSE*; *Eleanor Rigby*; *Guernica* (Picasso)
- **first lines of poems** – The rain set early in tonight; When I am an old woman I shall wear purple
- **subjects** other than names or titles – drama; R&D (research and development); seeds; yachts.

These are followed by locators and possibly cross-references.

In an index of first lines of verse or of songs, the article at the beginning of a first line of verse is not usually transposed, and is also the **search term** in the index for that line.

¹⁰ See **Module C Section C2**: 'Multiple sequences' to find out when more than one index may be required.

Activity A2

Choosing a subject with which you are familiar, compare the indexes of several books, paying particular attention to the features listed below. You will need to look closely at each index to try and identify its strengths and weaknesses and how well it serves its intended users.

- Introductory statement. Is there one? Is it helpful?
- Multiple sequences. Is there more than one index sequence? Is this helpful for the user of the particular document?
- Term selection. Has the indexer chosen appropriate terms for the document and its users? Have all the indexable elements been included? Is it easy to find information you expect to be in the document, or that you already *know* to be in the document? Do all the entries lead you to useful information?
- Headings and subheadings. Is the index in direct or classified form? Are subheadings used to direct users to various aspects of a topic?
- Cross-references. Are these used effectively to guide you to preferred terms and related topics?

This activity should be repeated regularly. As you progress through the course, you will find yourself better able to judge the qualities of published indexes. Asking questions like these will also help you to assess your own work objectively.

Activity A3

Look at the indexes in some eBooks and the different ways they appear.

BROWNE, Glenda. Introduction to indexing: a collection of articles. North Carolina: Lulu. 2013

<http://www.lulu.com/shop/glenda-browne/introduction-to-indexing-a-collection-of-articles/ebook/product-20661076.html>

(Available as a free download from Lulu, together with pc e-Reader software)

This gives examples of two types of linked index.

Newcomers: a selection of articles for those new to indexing. Sheffield: Society of Indexers. 2012

<http://www.theindexer.org/categories/collections.htm>

This has a linked index. It is available on several platforms.

How are they better or worse than traditional print indexes?

Which styles work best?