

1. What is an index?

From Janette Walters' blurb for this program (rephrased from "indexing" to "the index"):

"[The] index ... is ... the important topics, facts, names, and concepts in a publication ... organiz[ed] ... in alphabetical order with page references, so that readers can easily locate the information they need."

++++

From Larry Bonura, *The Art of Indexing* (1994, John Wiley & Sons, Inc. ISBN 0-471-01449-4):

"An index is a retrieval device, an access aid."

"An index is a systematic topical analysis alphabetically arranged or arranged by function, command, procedure, or topic. It is a reader's most important map for locating information ... [and is] ... read in random-access style."

Bonura explains that an index differs from a table of contents and a concordance because an index "systematically analyzes the topics."

++++

From the American Society of Indexers Web site (www.asindexing.org/indfaq.shtml):

"an index is a systematic arrangement of entries designed to enable users to locate information in a document."

++++

From the *Chicago Manual of Style*, 14th Ed, Chapter 17, “Indexes”:

“A good index records every pertinent statement made within the body of the text. The subject matter and purpose of the book determine which statements are pertinent and which peripheral. An index should be considerably more than an expanded alphabetical table of contents. It should also be something other than a concordance of words and phrases.”

++++

From William Meisheid, *Successful Indexing With RoboHELP HTML Edition* (2001, Sageline Publishing, ISBN 0-9672570-4-2):

Search versus index

Search mechanisms are neither informative nor browseable. You must know what you are looking for to find it. While this may be fine for experts in the domain (field or subject matter), it is not realistic for those unfamiliar with the domain.

Indexes, on the other hand, are both informative and browseable. The user does not have to know anything about the information domain, since the index opens the information space and its concepts to the user....

What an index is not

An index is not

- **A concordance** – An alphabetical list of significant words in the help or online information system.
- **A list of topic titles** – Though topic titles are often entered as index keywords, this only produces an alpha-sorted table of contents (TOC). An index is neither an elaborate TOC nor an outline of the information's contents.
- **A commentary** – Indexes are not a place for the indexer's opinion on the text. Indexes are text-centric and exegetical (the analysis is *taken from* the text, not *read into* it).
- **An afterthought** – The index is not an accessory haphazardly added to the end of the information.

2. Why index?

Bonura:

The most obvious reason ... is because your customers want it ... Numerous surveys indicate that the most common complaints about technical documents are the lack of an index, or a poorly designed index ...

A thoughtful index increases the value of any [document]. It helps ensure that the [document] will be used often because ... [its] ... contents are accessible.

Adding value

A good index makes any [document] more valuable to all [users]:

- Researchers will not or cannot use a [document] that does not have one.
- Researchers find a good index helpful in evaluating a [publication].
- Readers prefer [publications] with good indexes, so they can find what interests them, or locate specific information.

If your document is going to be sold in a bookstore or purchased by a library, then consider:

- Libraries typically will not buy a book without an index.
- Bookstore browsers often make decisions about book purchases based on what they find, or do not find, in an index.
- Reviewers often praise the index as well as the book.

From Meisheid:

Why is indexing important?

Help and online information users want quick and easy access to information. Suppliers of information often think supplying a full text search mechanism or going a step further and integrating a natural language search will meet the users' needs. That is not necessarily correct. ...

[N]atural language query processes will help many people where full text search has failed them. However, there are two things it can never replace, due to the need for the user to formulate a question: the informative presentation of possibilities and the browseability of an index.

(Meisheid later explicates “the informative presentation of possibilities”: “Properly constructed, with *See* and *See also* cross-references, the index is also instructive, explaining to the user the language of the domain.”)

From Kurt Ament, *Indexing: A Nuts-and-Bolts Guide for Technical Writers* (2001, William Andrew Publishing, ISBN 0-8155-1481-6):

**Usable indexes increase profits...[because]
Usable indexes improve documentation...
Usable documentation improves products...[and]
Usable products sell...**

3. What makes a good index?

A good index is:

- **Accurate**
- **Complete**
- **Concise**
- **Cross-referenced**
- **Logical**
- **Reader-appropriate**
- **Reliable**
- **Usable (and reusable) --from Maria Coughlin**

From Meisheid:

No such thing as a complete index...

No such thing as a perfect index...

Measurable criteria:

A good index balances recall (retrieval of as much relevant material as possible) and precision (retrieval of only the most relevant items). While it is hard to measure how well any index meets that standard, it is possible to evaluate an index based on certain criteria.

A good index should be

- **Encompassing** – References for all similar concepts scattered throughout the information space are gathered into the index.
- **Accurate** – Each item references a relevant topic or Web page.
- **Comprehensive** – Every relevant term or concept is in the index.
- **Concise** – Items and concepts are references by well-chosen terms.
- **Parallel and consistent** – Terms are grammatically comparable and similarly represented.
- **Double posted** – Entries are inverted or double posted as main entries as necessary to facilitate access.
- **Cross-referenced** – The index is adequately cross-referenced using *See* and *See also* references.
- **Text-centered** – The index serves the text of the information space and not the prejudices of the indexer.

From Bonura:

Five criteria for a good index

A good index has the following characteristics:

- Accuracy
- Depth of indexing
- Conciseness
- Cross-referencing
- Logical headings

Plain-language rules for indexing

- **If you pick it up, pick it up.** – Do Mi Stauber
- **It's better to be consistent than to be right.** – Dick Evans
- **You can have fast, cheap, good – pick two.** – Dick Evans
- **Know thy audience.** – Bill Meisheid

4. What goes in the index?

GLOSSARY

acronyms If you pronounce an abbreviation as a word (“ASCII”), it’s an acronym.

body The body of the text is the narrative “stuff” that isn’t in titles, headings, diagrams, tables, or boxes. It includes lists and examples.

concept The topic of an entry. It does not necessarily use exactly the same word, words, or phrasing as the text. *See also keywords.*

cross-references

- A *See* reference leads from a term not used as an index entry to the synonym or alternate term that is used (the target term).
- A *See also* reference leads to related terms (information) under another heading (target term).
- A **missing entry** is a cross-reference to a target term that doesn’t exist. An absolute No-No.

entry An entry is a topic plus a locator. There are *levels* of entries:

- a **main entry** is the primary (first-level) entry.
- **subentries** are the secondary (tertiary, etc.) levels that fall under the main entry

headings In the text, a heading is a word or phrase (usually in a different type size and font from the narrative text and set off with space above and below it) that divides one part of the narrative from another. In an index, “heading” is a synonym for “entry.”

keywords Keywords are significant *words* that are prime candidates for index entries. They differ from topics in that every keyword has a topic, but not every topic has a keyword. While a concordance is a list of words, an index uses words, but only insofar as they serve as concepts (topics) and modifiers. Confused? Ignore this definition entirely and think Keyword = Entry.

locator A locator indicates where the concept is to be found. It may be a page number (also called a page reference), a tag/URL, or a compound of section + paragraph number or volume + issue + page numbers.

topic The topic is the subject of the main index entry.

What to index – Technical Publications

- ✓ **Acronyms**
 - ✓ **Alternate names and common synonyms (?)**
 - ✓ **Command names**
 - ✓ **Error conditions and messages**
 - ✓ **Examples (?)**
 - ✓ **Figures**
 - ✓ **Glossary terms**
 - ✓ **Introductory information (?)**
 - ✓ **Keyboard keys and shortcuts**
 - ✓ **Measurements**
 - ✓ **Menu options**
 - ✓ **Overviews (?)**
 - ✓ **Proper names**
 - ✓ **Screen selections**
 - ✓ **Restrictions**
 - ✓ **Special characters or symbols**
 - ✓ **System messages**
 - ✓ **Tables**
 - ✓ **Tasks**
 - ✓ **Tools**
- (?) = maybe (it depends)**

5. Before you begin

As you follow the steps outlined below, always keep in mind that you should index topics in the same manner in which the reader/user will attempt to find them. As Ament says: “Walk in the shoes of your users. Look at the world from their perspective. Speak *their* language. Localize your index to the idiosyncrasies of your users.”

Also, don't try to index the whole document at once. Approach the document by chapter or section and work through them one at a time, completing each one as you go.

6. The indexing process, step by step

Step 1. IDENTIFY YOUR USERS

If you're the author of the document, you already know its organization, orientation, and content, and you are probably prepared to determine suitable indexing topics. If you're not the author, you need to determine who will be using your index. If you can ask the author in person, because she sits in the next cubicle, then ask her: Who is your audience? Programmers? Tech support? Engineers? End-users? Students? If the document has a preface or introduction, you can usually glean clues from reading it. (If you're not the author, this is also a good time to glance through the text, noting its organization and content, particularly any boxed materials, bold-faced or italicized terms, lists, glossaries or vocabulary terms, summaries, or “what you've accomplished”-type lists.)

Step 2. INDEX THE CHAPTER TITLES

Analysis: Is the chapter title descriptive (narrative) or prescriptive (procedural)? Descriptive titles answer the classic “who-what-where-when-why” quintet.

Prescriptive titles answer the solo “how.”

Method: To create an index entry for a descriptive title, choose the prominent answer(s) to the quintet (e.g., mice, men, monkeys). To create an entry for a procedure (task), choose a term that describes the procedure (e.g., installing, exiting, scratching). For the locator, enter the page or page range that covers the topic (e.g., 3 or 3-10). Index all the chapter titles at once, so that you get an overview of the whole document.

Step 3. INDEX TASKS (PROCEDURES)

Analysis: Tasks described in the main body of the text are usually easy to locate. They are often contained in headings, they indicate an action (e.g., “strangling your cat”), and they contain instructions. You can index the tasks in each chapter as you work your way through it, or you can create index entries for all the procedures in the text at once. Don’t forget tasks discussed in appendices.

Method: Create an entry for the task. If the discussion of the task is longer than one page, use a page range for the locator.

Step 4. FIND TOPICS

Analysis: Topics in the text also answer the “who-what-when-where-why” quintet. They may be contained in headings, they may be printed in boldface or italic type, and they may be ideas in words, phrases, paragraphs, examples, table titles, table contents, diagrams, or glossaries. *Selecting topics is the most challenging part of creating an index.* Here are some hints for analyzing topics:

- **Names** are usually topics (including product names, people’s names, place names, and company names) if they are discussed meaningfully, not given as examples or mentioned in passing. (Ask yourself: If the user comes to this page, will he learn anything substantial or relevant about “X”?) Drugs have generic names and brand names. They are usually entered under their generic name (“aspirin”) with the brand name cross-referenced back to the generic (“Bayer. *See* aspirin”). [NOTE: If a person, product, company, or drug is the primary topic of the whole document, you will probably not enter it in the index! That’s because the primary topic is understood to be the context for all your index entries.] Cross-reference all trademarked names (“WinWord. *See* Microsoft Word”).
- This is a good point for our discussion of **capitalization**. It’s important that you always capitalize proper names, including product brand names, in the index. Currently, the trend in indexing

is to leave all other terms lower-cased, even when they are the main entry.

- **Definitions** are easily identified as topics. Definitions can occur in the body of the text or in the glossary. Although sometimes the glossary is not indexed, all definitions given in the body of the text should be indexed, because definitions are keys to understanding the document.
- **Acronyms** and **abbreviations** are shortened forms of names, words, or phrases. Familiar (common) abbreviations usually can serve as main entries *because they're more familiar to the user than the terms they abbreviate* (e.g., DNA, IBM, URL, XML). Uncommon abbreviations or acronyms or those that stand for more than one thing are usually cross-referenced to their fully spelled-out form (“UMB. *See* University of Maryland at Baltimore; Untitled memory block”). Be *consistent* in your approach to abbreviations. Most often, your procedure will be to use the acronyms and abbreviations as the main entries and to cross-reference their fully spelled-out versions to the acronym/abbreviation as the target (“Extensible Markup Language. *See* XML”).
- **Warnings, notes, error conditions and messages, and system messages** are fairly obvious topics. For example: if a printer manual says “!WARNING: Spilling liquids on the printer will cause an electrical hazard” you should index that topic, probably under both the term “electrical hazard” and the category “warnings,” and maybe also under

“safety.” Can you think of another topic you might also enter it under? (Not “stupidity.”)

- **Special characters and symbols** present an interesting challenge. They are clearly topics, but where do they go in the index? At the top, under their actual symbol (“™, use of, 20”)? In the text, under their spelled-out form (“trademark (™), use of, 20”). Or both places? Where would you look for them?
- **Command names, menus, screen selections, tools, and keyboard keys and shortcuts** are other obvious topics. But don’t make them into hierarchical lists, because that puts a non-intuitive layer between the user and the index entry. For example, does the reader know that “Edit” is both a command and a menu? A user interested in editing should be given an indication of both conditions in one stop.

Bad:

Commands

Add, 10

Edit, 15

Save, 20

Menus

Document, 14

Edit, 15

Tools, 22

Good:

Add command, 10

Document menu, 14

Edit command, 15

Edit menu, 15

Save command, 20

Tools menu, 22

BUT you can “nest” commands as subentries under their menu, because the user interested in editing can browse the possibilities. In the CINDEX™ manual, the entry looks like:

Edit menu, 248-256

Undo, 248

Cut, 248

Copy, 248

Paste, 248

Clear, 248

Select All, 248

New Record, 249

Edit Record, 249

Duplicate, 249

Deleted, 249

Labeled, 250

Find, 250-252

Replace, 252-253, 252*f*

New Group, 253

Save Group, 254

New Abbreviation, 254

Preferences, 254-256, 254*f*

What order are these subentries in? What does *f* mean?

- **Figures (diagrams)** and **tables** usually state their topic clearly in their legend (caption) or title. Some indexers put *f* or *t* after the locator to indicate that the index entry is found in the figure or table, respectively.

- **Restrictions** are not as easily identified as topics, but they are. Look for clues like rules, cautions, notes, or default values and options. In other words, things you can do or change, but with (possibly dire) consequences.

An indexing manual, for example, might have the entry:

Sorting, 137-156

 rules of, overriding, 149-156

The subentry here is a restriction, and if you go to pages 149-156, you'll find all the consequences of overriding the sort rules.

- **Alternate names and common synonyms, examples, introductory information, and overviews** *may* be topics, depending on your users. Remember, know your users and you'll know whether they'll need to find these topics.

Step 5. ENTER/EDIT YOUR TOPICS

Method: Indexers call this “casting entries.” (My dictionary’s definitions of “cast” include “**1.**To throw, hurl. **2.** To shed, molt. ... **9.** To give birth to prematurely. **10.** To cause (hounds) to scatter and circle in search of lost scent.” These are nicely colorful versions of what *I* do when I cast entries, but the accepted definition is probably “**13.** To give form to; arrange.”) You might think of it as the phrasing or syntax of your index entries. There are myriads of rules about casting entries, But here’s a fast overview:

- Emphasize the major noun (or subject or keyword) of the entry by putting it as close to the left margin as possible. (Users scan the left margins of indexes.) For example, a pet owner’s manual says: “CAUTION: Feeding corn to a snake can kill it.” Using “feeding corn” is not an effective entry. The keyword is *corn*, so the entry should be “corn, feeding to snake.”
- Make as many entries as necessary to help the user find the information. In our example above, we would also invert the entry to read “snake, feeding corn to.”
- Later on, the pet owner’s manual also contains the information that corn gives dogs a bellyache. Not only does this sentence have two keywords, *corn* and *dog*, but also it contains information about corn that is of the same general class as that given for snakes/corn. So, reanalysis of your snake entry in light of your dog entry indicates that it’s not the pouring of corn down the snake’s throat (*feeding*) that’s the topic, it’s the adverse effects of corn. So you edit your snake/corn and enter your dog/corn and get:

corn

adverse effects of

in dogs, 45

in snakes, 42

dogs

adverse effects of corn in, 45

snakes

adverse effects of corn in, 42

- By now you've noticed that index entries do not necessarily use natural grammatical order. Whereas the natural order is "knitting sweaters for penguins," the index order is "sweaters, for penguins, knitting," and "penguins, sweaters for, knitting."
- In the corn/snake and corn/dog examples above, the entries were also inverted, to make both instances of the relationship corn/animal available under both "corn" and the specific animal. Another use of inversion is in tasks, where both the task and its object are inverted: "files, editing" and "editing, files." However, do not invert "married terms." Married terms are adjective + noun compounds like "hot key" or "hard space" and noun + noun compounds like "quotation marks" or "disk space." Also, don't use an adjective alone as a main entry. The entries

binary
 language, 14
 operators, 11

should be entered as

binary language, 14
 binary operators, 11

- Beware of placing entries into false categories (Bonura calls this *factoring*). If factored, the potential entries "pasture fencing" and "fencing stolen goods" (not likely to appear in the same document, I admit) would come out:

fencing
 of pastures
 of stolen goods

- Locators are usually separated from the last word of the entry by a comma or extra space (or sometimes by a colon). Multiple locators are separated from each other by commas. If the locator covers a range, say pages 46 to 48, it is entered as 46-48, with a hyphen or a dash indicating the range. If the locators in a range already contain hyphens (if they indicate a section and a paragraph, for example) the word “to” is used to indicate the range: for example, 10-22 to 10-24, not 10-22–10-24 or, worse, 10-22–24. If discussion of a topic begins on page 3, is broken by another topic and then is resumed on page 4, the locator could be either “3-4” or “3, 4” as long as the indexer is consistent and always uses the same format for instances of this type.

Step 6. FORMATTING THE INDEX

Analysis: Formatting is how the index is sorted and displayed and what punctuation is used – in other words, what the index looks like.

Method: If the publisher of your document has preferences (alphabetical sorting by word or by letter, indented index or run-in – paragraph – style, cross-references located at top or bottom of entry), you’ll probably receive a style sheet to follow. If nobody has any idea what to do, pick an index you like and follow it. NOTE: Until you get lots of indexing experience, stay away from run-in indexes. It’s much more difficult to cast entries for run-in indexes, and they’re best avoided if the index requires more than one level of subentry. ANOTHER NOTE: The best way to format an index is to use dedicated indexing software!

6. Tools and resources

Organizations and SIGs

American Society of Indexers

10200 West 44th Avenue, Suite 304

Wheat Ridge, CO 80033

Voice: 303-463-2887

Fax: 303-422-8894

Email: info@asindexing.org

ASI has several SIGs available to members.

STC has an active indexing SIG. SIG members receive “A to Z,” the SIG’s newsletter, which is published three times a year, and they have a listserv. SIG members can subscribe by sending e-mail with the contents “subscribe stcsig-1 [your e-mail address]” to lyris@lists.stc.org.

Indexing software

CINDEX™

Contact information:

Indexing Research

100 Allens Creek Road

Rochester, NY 14618

Voice: 716-461-5530

Fax: 716-442-3924

email: flennie@indexres.com

web: www.indexres.com

Demo version available as download.

IXGen

Contact information:

Web: <http://home.pacifier.com/~franks/index.html>

Excellently reviewed by Anne C. Barrett in the September 2001 issue of "A to Z."

Macrex (I have no information)

SKY Index™

Contact information:

SKY Software

350 Montgomery Circle

Stephens City, VA 22655

Toll-free: 800-776-0137

Local: 540-869-6581

Email: <mailto:kamm@sky-software.com>

web: www.sky-software.com

Demo version available as download.

Leverage Technologies (LevTech) supplies training, installation, and support for CINDEX™. LevTech also supplies utilities for CINDEX™ that validate index style and help prepare Web indexes.

Contact information:

Leverage Technologies, Inc.

9519 Greystone Parkway

Cleveland, OH 44141-2939

Toll-free: 888-838-1203; Local/fax: 440-838-1203

e-mail: info@LevTechInc.com

web: www.LevTechinc.com

Publications

Kurt Ament, *Indexing: A Nuts-and-Bolts Guide for Technical Writers* (2001, William Andrew Publishing, ISBN 0-8155-1481-6) \$40+ (Amazon or the publisher)

Larry Bonura, *The Art of Indexing* (1994, John Wiley & Sons, Inc. ISBN 0-471-01449-4) \$39.99 on Amazon

Linda K. Fetters, *Handbook of Indexing Techniques : A Guide for Beginning Indexers* (1999, publisher and ISBN unknown to MC), \$20 on Amazon

William Meisheid, *Successful Indexing With RoboHELP HTML Edition* (2001, Sageline Publishing, ISBN 0-9672570-4-2) and *Teach Yourself Indexing for RoboHELP Classic* (2000, Sageline Publishing, ISBN 0-9672570-3-4), \$55 each. Sageline Publishing's contact information:

Sageline Publishing Bootstrap Books TM
502 Oella Avenue, Ellicott City, MD 21043
Voice: 410-465-2040
web: www.sageline.com

Nancy C. Mulvany, *Indexing Books* (1994, The University of Chicago Press, ISBN 0-226-55014-1), \$32, but \$22.40 on Amazon

The American Society of Indexers offers several publications on indexing through Information Today,

Inc. Descriptions of the publications are available on the ASI website or from Information Today (<http://www.infotoday.com>).

Indexing listserves

The STC indexing SIG has its own listserve (see above, under SIGs).

To subscribe to Index-L, a general indexing chat list, send e-mail with the contents “subscribe INDEX-L [your name]” to lyris@listserv.unc.edu.

IndexPeers

Indexers volunteer to review each other's edited indexes. To subscribe, go to groups.yahoo.com/group/IndexPeers. To contact the listowner, email index@teleport.com.