Hands-on Indexing

Winfield Swanson

Computers can simplify the indexer's life by a huge degree. At least it's tidier—I will never again have to prepare to compile an encyclopedic index for an 800-page book by buying six bread pans to hold the index cards, nor wonder what to do with all the half-used cards.

Computers are excellent for compiling indexes that require no thought, for instance, indexes of authors' names. Computer-generated indexes for the subject matter of books and volumes of journals are also an option, but a poor one. On closer inspection, you will find that the machine does exactly what it was instructed to do. Namely, it takes every word you tell it to every time that word occurs, regardless of whether the word is discussed or only mentioned and regardless of context—thoughtlessly, in fact. For instance, Black might be Andrew Black, Black Skull (famous fossil), Black Hills, or black people; ADA could be the American Dental Association, adenosine deaminase, or (acetamidol)iminodiacetic acid; and so forth. Worse, the text might simply say that ADA is not a member of some group under discussion, or the entry might be one in a long list and never be mentioned again. To complete this index, the indexer must look up each entry to determine whether it belongs in the index, thereby losing all the time—and more—saved in keying in topics. Subheads are also a problem in that the computer system used might not have been able to make them, so each topic appears in the index as equally important. Never agree to "just look over" a computer-generated index. It takes days.

A number of software packages (such as CINDEX and MACREX) offer immediate sorting to several levels. Besides their expense, their disadvantage is that if you have committed a typo or can't remember the term you used, the computer has already sorted the entry the way you typed it and you might never find it again. Most people do not need this level of complexity. If your goal is a humbler one- or two-level index for a book or a volume of a journal or magazine, any word processor will suffice.

***Winfield Swanson** is a publishing consultant in Washington, DC.*

The Mechanics

In computerspeak, *alphabetize = sort*. The sort command is at the bottom of the pull-down menu labeled "Table" on the Mac 7200 with Microsoft Word 98 software I am now using; in an earlier version (Word 5) it was under "Tools". Whatever word-processing program you use, you can sort by paragraphs or fields, in ascending or descending order. You want paragraphs (by definition, a paragraph is a body of text that starts and ends with a hard return) in ascending order (A-Z). Alphabetizing a list of words from A to Z seems like a simple, straightforward task. However, you should be aware of, and allow time for, some computer-generated curiosities, including the following. Here is a computer-sorted list:

blood donors, 475, 478
blood, 110, 121 (macaques), 124, 392, 556, 785, 841
blood-brain barrier, 292

Note that if a word is hyphenated (*blood-brain barrier*), it will appear after entries that have the same word not hyphenated, regardless of spelling. If an entry contains more than one word, it will come before an entry that has only the first of those words (*blood donors* comes before *blood*). The human-sorted list would read like this:

blood, 110, 121 (macaques), 124, 392, 556, 785, 841
blood-brain barrier, 292
blood donors, 475, 478

Furthermore, if an entry begins with a quotation mark, it might be sent to the very bottom of the index. Computers sort word by word and not letter by letter (current fashion—like logic—dictates letter by letter).

computer-sorted, word by word:
environment and the law, 35
Environment Department (ENV), 284, 299
environmental law, 35
environment-related documents, 98-114
"Environment on the Moon", 260

human-sorted, letter by letter:
environmental law, 35
environment and the law, 35
Environment Department (ENV), 284, 299
"Environment on the Moon", 260
environment-related documents, 98-114

Computers used to (and maybe some still do) sort capitalized words before lower-case words.

computer-sorted:
Everngam, Ray, 1999 (22/2):37
embargoes, 1999 (22/3):82
ethics, guidelines for communication of science, 1999 (22/3):79-80
see also integrity
e-mail etiquette, 1999 (22/5):165
ethnic groups, terminology, 1999 (22/5):166-7

human-sorted:
e-mail etiquette, 1999 (22/5):165
embargoes, 1999 (22/3):82
ethics, guidelines for communication of science, 1999 (22/3):79-80
see also integrity
ethnic groups, terminology, 1999 (22/5):166-7
Everngam, Ray, 1999 (22/2):37

These curiosities are gradually disappearing. (If you know how to make a computer sort letter by letter, please tell me.)

The first step is to simply key in (indexers of a certain age will want to type) the entry (also known as the heading) and the page numbers. If you want to add a subhead, be sure not to separate it from its main head with a hard return, because the computer will alphabetize (sort) entries after a hard return as though they are all of equal weight (so your subheads will become main heads).

Depending on the situation, it may be useful to keep two copies of your index: one in the order in which you originally listed the entries (the first iteration) and one alphabetized (sorted). This form of insurance will enable you to change page numbers more easily if the layout changes (a situation to be avoided if at all possible); or you can more easily find a missed fact that you later want to include (also to be avoided if at all possible). For example:

First draft (original listing of entries):

copyright, 7, 8 duration, 8
copyright, 15-56
sculpture, 18-19
audiovisual works, 19-20
architectural works, 20-21
compilations, 21-23
art, arrangement, 22
exclusions, 23-24
facts, 23 historical, 23-24
copyright, before January 1, 1978, 24
art, visual, 25
copyright, after March 1, 1989, 25
art, moral rights, 31-32
art, removal, 32
art, modifications, 32
copyright, duration, 32-41
audiovisual works, 56
sculpture, as trademark, 71-72
European Union, 161

Second draft (entries sorted by computer):

architectural works, 20-21
art, arrangement, 22
art, modifications, 32
art, moral rights, 31-32
art, removal, 32
art, visual, 25
audiovisual works, 19-20
audiovisual works, 56
compilations, 21-23
copyright, 15-56
copyright, 7, 8 duration, 8
copyright, after March 1, 1989, 25
copyright, before January 1, 1978, 24
copyright, duration, 32-41
European Union, 161
exclusions, 23-24
facts, 23 historical, 23-24
sculpture, 18-19
sculpture, as trademark, 71-72

Third draft (subheadings manually separated from main headings, manually indented, and alphabetized by computer and you):

architectural works, 20-21
art
  arrangement, 22
  modifications, 32
  moral rights, 31-32
  removal, 32
  visual, 25
audiovisual works, 19-20, 56

compilations, 21-23
copyright, 7, 8, 15-56
  after March 1, 1989, 25
  before January 1, 1978, 24
  duration, 8, 32-41

European Union, 161
exclusions, 23-24

facts, 23
  historical, 23-24

sculpture, 18-19
  as trademark, 71-72

Format

Before you begin, be sure you know who will be using the index and what format the publisher wants. In our simplified world, indexes with a main head and one subhead are the norm—entries that can be understood at a glance without giving unnecessary information. Indexes with subheads and sub-subheads and sub-sub-subheads have the old-fashioned charm of the days when people expected to pore over books.

CSE style recommends using lowercase for all entries except proper nouns, using indented subheads, and not repeating digits if space is a concern; for example:

burials and graves
  Christianity and, 659
  collective, 40
  cremation, 591, 614-25, 659, 754
  "giants' tombs," 534
  religion and, 591, 592-3
  social organization and, 661
  wood artifacts, 759
  see also bog bodies; megalithic monuments; pyramids; reburial;urnfields

fission-track dating, 169, 238, 387
  see also dendrochronology; potassium-argon dating; radiocarbon dating

But many publishers retain the custom of capitalizing main heads and using lowercase for subheads. A few publishers instruct indexers to run in the subheads in paragraph form and separate the main head from the subheads with a colon. The subheads with their page numbers are separated from each other with a semicolon like this:

burials and graves: and Christianity, 659; collective, 40; and cremation, 591, 634-25, 659, 754; "giants' tombs," 534; and religion, 591, 592-3; and social organization, 661; wood artifacts, 759. *See also* bog bodies; megalithic monuments; pyramids; reburial; urnfields

fission-track dating, 169, 238, 387. *See also* dendrochronology; potassium-argon dating; radiocarbon dating

Page numbers are usually separated from text and each other with commas, but page ranges offer several possibilities. The numbers may be separated with a hyphen, but an en dash is more easily read. Some styles

recommend stating each number (152-158, 101-102), some recommend stating only the different number (152-8, 101-2), and some recommend following the usage common in speech (152-58 but 101-102). Of course, each comes with its own set of rules. Read the publisher's instructions before you begin.

Newsletters add another wrinkle in that many do not have consecutive page numbers for successive issues. That is, each issue begins with page 1, so to distinguish among the page 1s, you must add the month and maybe the volume number and the year. Here's one solution:

nitroglycerin
metered-dose-Oct '85:36; May '86:32
packaging and labeling-May '85:39
nuclear catastrophe
Chernobyl-Apr '89:38-39; Oct '92:18
Three Mile Island-Oct '92:17-18

The End

Managers trying to check for accuracy sometimes look up page numbers to be sure that a term is actually cited on the page shown. That is fine. Some indexers do get the wrong page numbers, but not often.

A good index is not only a list of alphabetized words and correct page numbers. The real test of an index is in

- the topics that were (or were not) selected
- how the topics were (or were not) cross-referenced
- the terms selected for the audience (will the likely reader go to *hematology* or to *blood*? to *cancer* or to *non-Hodgkin's lymphoma*, *melanoma*, or *breast cancer*?)
- the concepts indexed or cross-referenced to the appropriate term

Clearly, tedium is still a byword of the whole process. I like to think of indexing as the manual labor of the editorial biz—low prestige but relatively high pay, if your mind works that way (and, if your mind does work that way, it's a cut above crossword puzzles). How much you can get the computer to do for you depends directly on your word-processing skills. None of this is glamorous or prestigious. Nevertheless, creativity, thought, knowledge, and experience come into play. Those are things that a machine doesn't have, and that is why publishing (both electronic and print) still has a place for the human indexer.

What's New

a new job?
a prize?
an award?
publication of a book?
a degree or certificate?
recent retirement?

If you have something new in your professional life, we invite you to share the news with readers of *Science Editor*. Please supply a brief announcement to Barbara Gastel, Editor, *Science Editor*, Department of Journalism, 230 Reed McDonald Building, Texas A&M University, College Station TX 77843-4111, fax 979-845-6887,
e-mail *b-gastel@tamu.edu*.