

PERSPECTIVES ON . . .

Indexing

EDITOR

Lois F. Lunin

922 24th Street, NW, Washington, DC 20037-2229, and the Department of Radiology, Cornell University Medical College, New York, New York 10021

GUEST EDITOR

Raya Fidel

Graduate School of Library and Information Science, University of Washington, Seattle, Washington 98195

Introduction and Overview

The design of an information system assumes that the system has one or more users who will search for information, using keys appropriate to that system. Although the technologies to process information have become infinitely more sophisticated in the latter part of this century, recognition that indexing is critical to the efficient operation of systems is not new. Early- to mid-20th century systems concentrated on various techniques such as metalanguages and codes to characterize documents and to make them amenable to systems then available, such as punched cards.

Keys in one form or another are still needed, and new kinds are called for because technologies today make it possible to process and store many more forms of information than text or numeric data. These new forms—such as images; photographs; video, still and motion; multimedia; and sound—pose new challenges to indexing. Academicians as well as several commercial companies have recognized many of these challenges and are conducting research or offering systems that index automatically, use bar code, OCR/ICR (Optical Character Recognition/Intelligent Character Recognition), mark-sense boxes, and use neural net processing, pattern recognition, and, of course, full text with varying combinations of human and automated indexing.

This issue of *Perspectives* takes a first step to look at the intellectual or cognitive aspects of indexes of text and images from which researchers and system designers may derive new principles or refine those in use. It explores ways to think about the new needs and new capabilities. The rapidly developing Internet and the increasing number of databases worldwide emphasize the importance of good indexing as well as the growing complexity and difficulties in that analysis. Examples of these complexities are systems to index e-mail messages that administrations must by law save, and authentication of documents in digital form to verify the original or indicate a modification. While technologies ranging from smart cards to supercomputers and software with gophers appear to receive the most attention at conferences and in publications, the heart—indeed the soul—of many systems is the indexing, whether automated or human.

A recurring motif in this issue of *Perspectives* is the need to consider user needs in indexing theory and prac-

tice. The first article “User-Centered Indexing” by Raya Fidel sets the scene. It briefly describes the place of user-centered approaches in research about indexing and shows the position of each article published in this issue within a research framework. Using a different angle, Jessica Milstead creates a new framework in her article, “Needs for Research in Indexing,” when she answers what research practicing indexers need to support their work. She uncovers issues that require scientific investigations in the areas of cognitive processes of both the index creator and user, vocabulary control, and structure and layout of indexes.

Two articles address specific aspects of indexing. Sara Shatford Layne focuses on groupings of images in her article, “Some Issues in the Indexing of Images.” She first defines four distinct categories of image attributes—biographical, subject, exemplified, and relationship—and then explains why it is important to provide access to useful groupings of images, rather than to individual images, and the issues involved in creating such groupings. In his article, “Indexing and Retrieval Performance: The Logical Evidence,” Dagobert Soergel addresses a topic whose popularity faded at the end of the 1970s: How does indexing affect retrieval performance? Pointing to the complexity involved and the difficulty in achieving reliable results in empirical studies, he first defines indexing characteristics that might affect performance and then examines in detail, using logical evidence, their effect on retrieval.

Three articles examine indexing of specific materials. “Access to Nonbook Materials: The Limits of Subject Indexing for Visual and Aural Languages” is the title of Elaine Svenonius’s contribution. She uses an aboutness model of indexing to examine some materials with non-textual information. She concludes that there are instances that defy subject indexing, not because of the nature of the medium *per se*, but because it is being used for nondocumentary purposes. Helen Tibbo’s article is entitled “Indexing for the Humanities.” She demonstrates that there are two major reasons to require new controlled vocabularies and indexing framework for the humanities: (a) humanists use a wide variety of textual, graphic, and aural materials in their research; and (b) indexing and surrogating models from the sciences are not adequate to meet the information access needs of humanists.

The seventh article, by Hans Wellisch, examines the

difference between indexing different types of materials. In "Book and Periodical Indexing," Wellisch develops a typology of units of verbal texts, starting from books written by a single author to articles in periodicals. He discusses problems in indexing each type, and at the same time emphasizes that indexing of *all* types must fulfill certain basic functions. These functions are prescribed by indexing standards. Concluding this issue is James Anderson's article "Standards for Indexing: Revising the American National Standard Guidelines Z39.4." He describes the nature of standards in general, and of the current draft of the *Standard Guidelines*, and the practical and intellectual work that was required for the creation of the new standard.

Although this issue covers several aspects of indexing, including reference to automated indexing, space has limited its scope to these articles. It is planned to include other aspects, such as specifics of automated indexing, in a later issue.

Lois F. Lunin
Raya Fidel

About the Authors

James D. Anderson is Associate Dean and Professor in the School of Communication, Information, and Library Studies, Rutgers University, where he specializes in the design of textual databases and vocabulary management systems for information retrieval.

Raya Fidel, after working as a science librarian at the Hebrew University in Jerusalem, received her Ph.D. from the University of Maryland in 1982. Currently she is an Associate Professor at the University of Washington where she teaches courses about indexing and abstracting, conceptual database design, and information science. Because she maintains that design of information retrieval systems should be based on an understanding of users' information needs and behavior, her research has been focused on online searching behavior. Her work has been published in various professional journals.

Sara Shatford Layne has a B.A. and M.F.A. from Stanford University and an M.L.S. from the University of California, Los Angeles, where she is a doctoral candidate in the Graduate School of Library and Information Science. She worked as a theatrical costume designer for nine years and is now a cataloger at UCLA. She is also a consultant to the *Index of Medieval Medical Images* and to the Art Information Task Force.

Jessica L. Milstead is principal of the JELEM Company. She has undertaken assignments in indexing system design and thesaurus development for a varied clientele. She has also been responsible for production of bibliographic databases, and has carried out research on the state of the art of subject analysis.

Dagobert Soergel is Professor of Information Studies, University of Maryland, and an information systems consultant. He has authored *Organizing Information* (1985) which received the ASIS Best Book Award, *Indexing Languages and Thesauri* (1974), and numerous papers.

Elaine Svenonius has degrees in philosophy and library and information science. Her research has been and continues to be in the area of bibliographical control, including cataloging, classification, and indexing. Her particular focus has been on philosophical analysis and on the design of documentary languages in the automated environment.

Helen Tibbo, a faculty member of the School of Information and Library Science at the University of North Carolina at Chapel Hill, has done extensive research in the area of information retrieval for the humanities. Her dissertation from the University of Maryland, *Abstracting, Online Searching, and the Humanities*, won both the American Society for Information Science's Doctoral Forum Award and the Association for Library and Information Science Education's Special Research Award. This work is the basis of *Abstracting, Information Retrieval, and the Humanities: Providing Access to Historical Literature*, published in 1993 in the Association of College and Research Libraries Publications in Librarianship monograph series.

Hans H. Wellisch was born in Vienna, Austria and went to Israel in 1949, after he lived in Sweden for 10 years. In Israel, he was Head of the Information Center of Tahal, a civil engineering company specializing in water resources development. He also taught indexing and cataloging, and wrote the first Hebrew book on the management of special libraries. In 1969 he joined the faculty of the College of Library and Information Services of the University of Maryland, from which he received his PhD. Since 1987 he has been Professor Emeritus. His interests center on all aspects of bibliographic control. In 1979, Dr. Wellisch received the first H. W. Wilson Company Award for Excellence in Indexing for the index he compiled for his book *The Conversion of Scripts*. His most recent book is *Indexing from A to Z*. He is active in the work of NISO on standards for information work, most recently on thesaurus construction and indexing.