•ORGANIZING INFORMATION

Principles of Data Base and Retrieval Systems

LIBRARY AND INFORMATION SCIENCE

consulting Editors: *Harold Borko and* **G.** *Edward Evans*
GRADUATE SCHOOL OF LIBRARY SCIENCE
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Thomas H. Mott, /r., ***Susan Artandi, and Leny Struminger***
Introduction to PL/I Programming for Library and Information Science

Karen Sparck Jones and Martin Kay
Linguistics and Information Science

Manfred Kochen (Ed.)
Information for Action: From Knowledge to Wisdom

Harold Borko and Charles L. Bernier
Abstracting Concepts and Methods

F. W. Lancaster
Toward Paperless Information Systems

H. 5. ***Heaps***
Information Retrieval: Computational and Theoretical Aspects

Harold Borko and Charles L Bernier
Indexing Concepts and Methods

Gerald jahoda and Judith Schiek Braunagel
The Librarian and Reference Queries: A Systematic Approach

Charles H. Busha and Stephen P. Harter
Research Methods in Librarianship: Techniques and Interpretation

Diana M. Thomas, Ann T. Hinckley, and Elizabeth R. Eisenbach
The Effective Reference Librarian

James Cabeceiras
The Multimedia Library, Second Edition: Materials Selection and Use

C. ***Edward Evans***
Management Techniques for Librarians, Second Edition

Irene P. Codden (Ed.)
Library Technical Services: Operations and Management

Jessica L Milstead
Subject Access Systems: Alternatives in Design

Dagobert Soergel
Organizing Information: Principles of Data Base and Retrieval Systems

ORGANIZING INFORMATION

Principles of Data Base and Retrieval Systems

Dagobert Soergel

College of Library and Information Services
University of Maryland
College Park, Maryland

1985

ACADEMIC PRESS, INC.
(Harcourt Brace Jovanovich, Publishers)
Orlando San Diego New York London
Toronto Montreal Sydney Tokyo

ACADEMIC PRESS, INC.
Orlando, Florida 32887

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
24-28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Soergel, Dagobert.
Organizing information.

(Library and information science)
Bibliography: p.
Includes index.
1. Information storage and retrieval systems.
I. Title. II. Series.
Z699.S539 1985 025'.04 83-15741
ISBN 0-12-654260-0 (alk. paper)
ISBN 0-12-654261-9 (paperback)

PRINTED IN THE UNITED STATES OF AMERICA

85 86 87 88 9876 5 432 1

Contents

V
ISAR SYSTEMS OPERATION AND DESIGN

Preface

This book gives a theoretical base and a perspective for the analysis, design, and operation of information systems, particularly their information storage and retrieval (ISAR) component, whether mechanized or manual. Information systems deal with many types of entities: events, persons, documents, business transactions, museum objects, research projects, and technical parts, to name a few. Among the purposes they serve are to inform the public, to support managers, researchers, and engineers, and to provide a knowledge base for an artificial intelligence program. The principles discussed in this book apply to all these contexts. The book achieves this generality by drawing on ideas from two conceptually overlapping areas—data base management and the organization and use of knowledge in libraries—and by integrating these ideas into a coherent framework. The principles discussed apply to the design of new systems and, more importantly, to the analysis of existing systems in order to exploit their capabilities better, to circumvent their shortcomings, and to introduce modifications where feasible.

This book is intended for use in an introductory course on organizing and retrieving information (called, for example, "Introduction to the Organization of Information," "Introduction to Information Storage and Retrieval," or "Introduction to Information Science") offered in a school of library and information science, a business school, or a more broadly based information studies or information management program. Beyond that, it is meant to inspire, a modernization and integration of the library/information science curriculum. The book can be used for a broadly based course that teaches the general principles of ISAR and treats cataloging and reference service as specific areas of application. Such a course not only overcomes the artificial separation of cataloging and reference but also giyes students wide flexibility in choosing their first position and a sound base from which to strike out in many directions in the further development of their careers. It can be offered as ä package of designated sections of the cataloging and reference course, without changing any course numbers. Such a course can be extended to include students from business infor-

mation systems, journalism, and cognate areas: The theoretical base is common to all, but the application areas are different. This book is also suitable for self-study by practitioners who are looking for a sounder theoretical base for their daily work, A workbook with exercises and discussion of additional examples is in preparation; a draft is available from the author.

Information studies is a nascent field. It shows considerable confusion in its terminology, partly due to the lack of a prevalent conceptual framework: The same term is used for different concepts; different terms are used for the same concept. This book follows the terminology of major writers in the field but sometimes introduces a new term for a new concept or to replace an existing term that reflects a faulty concept analysis.

Throughout the book the development of ideas proceeds from the point of view that information specialists are professionals who cooperate with the user in determining information needs and who use their knowledge to design systems or to do searches to meet these needs, as opposed to merely looking for what the user thinks is needed.

Organization of the Book. Part I places information systems in context; it discusses the nature and structure of information and lays out the overall structure of an information system. Part II provides the basis for considering the design and use of ISAR systems in light of the objectives to be achieved, allowing for a discussion of the merits of design alternatives in Parts III through V. Part III deals with the logical representation of data and with structures for providing access to these data. It deals on a general level with the rules and conventions necessary in an ISAR system. Part IV focuses attention on subject retrieval (but many of the principles have more general application). It discusses the nature of index languages—terminological control, basic functions, and conceptual structure. Part V discusses indexing and searching and, in conclusion, testing and design of the system.

Acknowledgments. This book was developed from lectures given at the University of Maryland; the students' many questions forced me to sharpen my thinking, and their comments on successive versions of the manuscript were extremely useful. Norman Roberts, Harold Borko, and Raya Fidel all gave good advice, which, among other things, was instrumental in reducing this book to a manageable size. Jane Bergling and Marie Somers typed the manuscript many times over, somehow managing to interpret my scribbled revisions. My wife Lissa was always ready to examine and discuss ideas and to make suggestions concerning both content and form; she also spent hours editing and proofreading. I owe an intellectual debt to many in the field but above all to the pioneering spirit of Calvin Mooers.