Structured Information Files

RICHARD A. KAIMANN
Marquette University
Milwaukee, Wisconsin

A Wiley-Becker & Hayes Series Book

MELVILLE PUBLISHING COMPANY
Los Angeles, California

Copyright © 1973, by John Wiley & Sons, Inc.

Published by Melville Publishing Company,
a Division of John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

No part of this book may be reproduced by any means,
nor transmitted, nor translated into a machine language
without the written permission of the publisher.

Library of Congress Cataloging in Publication Data

Kaimann, Richard A
Structured information files.

(Information sciences series)
Bibliography: p.
1. Electronic digital computers–Programming.
2. Information storage and retrieval systems.
I. Title.

QA76.6.K34 029.7 73-8650
ISBN 0-471-45483-4

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Information Sciences Series

Editors

ROBERT M. HAYES
University of California
Los Angeles, California

JOSEPH BECKER
President
Becker and Hayes, Inc.

Consultant

CHARLES P. BOURNE
University of California
Berkeley, California

Joseph Becker and Robert M. Hayes:
INFORMATION STORAGE AND RETRIEVAL

Charles P. Bourne:
METHODS OF INFORMATION HANDLING

Harold Borko:
AUTOMATED LANGUAGE PROCESSING

Russell D. Archibald and Richard L. Villoria:
NETWORK-BASED MANAGEMENT SYSTEMS (PERT/CPM)

Launor F. Carter:
NATIONAL DOCUMENT-HANDLING SYSTEMS FOR SCIENCE AND TECHNOLOGY

Perry E. Rosove:
DEVELOPING COMPUTER-BASED INFORMATION SYSTEMS

F. W. Lancaster:
INFORMATION RETRIEVAL SYSTEMS

Ralph L. Bisco:
DATA BASES, COMPUTERS, AND THE SOCIAL SCIENCES

Charles T. Meadow
MAN-MACHINE COMMUNICATION

Gerald Jahoda:
INFORMATION STORAGE AND RETRIEVAL SYSTEMS FOR INDIVIDUAL RESEARCHERS

Robert M. Hayes and Joseph Becker:
HANDBOOK OF DATA PROCESSING FOR LIBRARIES

Allen Kent:
INFORMATION ANALYSIS AND RETRIEVAL

Robert S. Taylor:
THE MAKING OF A LIBRARY

Herman M. Weisman:
INFORMATION SYSTEMS, SERVICES, AND CENTERS

Jesse H. Shera:
THE FOUNDATIONS OF EDUCATION FOR LIBRARIANSHIP

Charles T. Meadow:
THE ANALYSIS OF INFORMATION SYSTEMS, Second Edition

Stanley J. Swihart and Beryl F. Hefley:
COMPUTER SYSTEMS IN THE LIBRARY

F. W. Lancaster and E. G. Fayen
INFORMATION RETRIEVAL ON-LINE

Richard A. Kaimann:
STRUCTURED INFORMATION FILES

Thelma Freides:
LITERATURE AND BIBLIOGRAPHY OF THE SOCIAL SCIENCES

Dagobert Soergel:
INDEXING LANGUAGES AND THESAURI: CONSTRUCTION AND MAINTENANCE

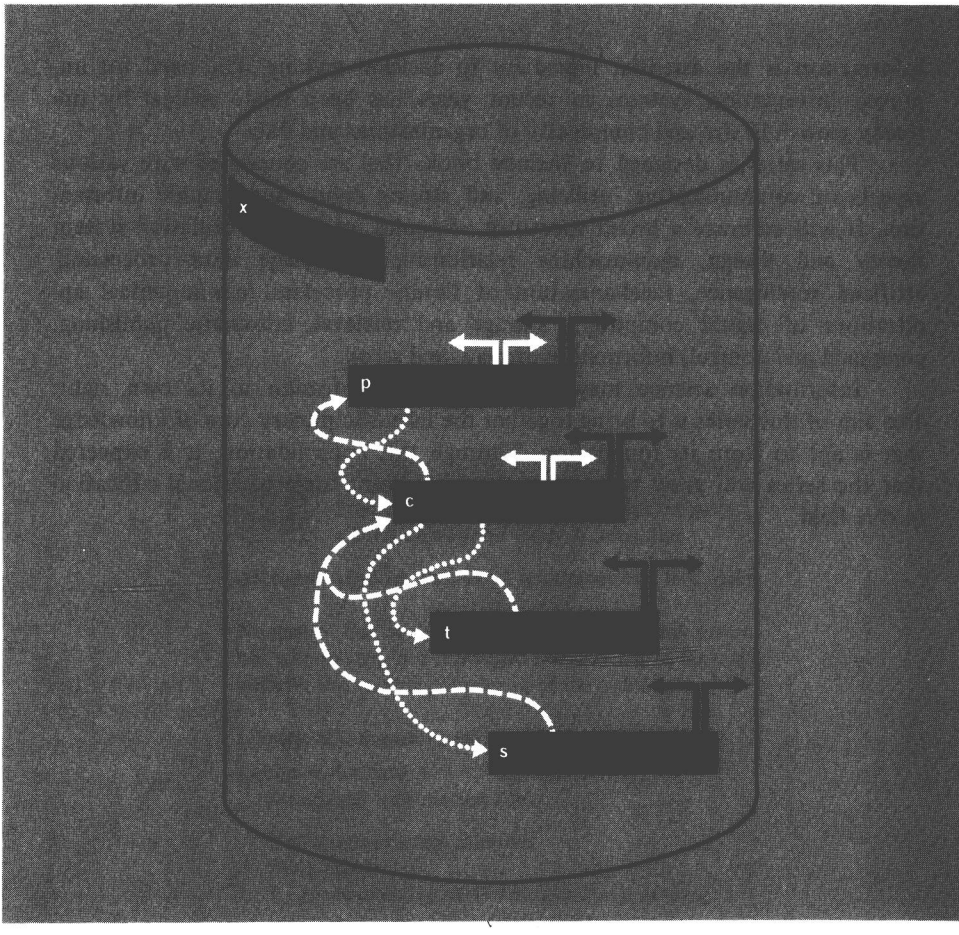
此为试读,需要完整PDF请访问: www.ertongbook.com

Information Sciences Series

Information is the essential ingredient in decision making. The need for improved information systems in recent years has been made critical by the steady growth in size and complexity of organizations and data.

This series is designed to include books that are concerned with various aspects of communicating, utilizing, and storing digital and graphic information. It will embrace a broad spectrum of topics, such as information system theory and design, man-machine relationships, language data processing, artificial intelligence, mechanization of library processes, nonnumerical applications of digital computers, storage and retrieval, automatic publishing, command and control, information display, and so on.

Information science may someday be a profession in its own right. The aim of this series is to bring together the interdisciplinary core of knowledge that is apt to form its foundation. Through this consolidation, it is expected that the series will grow to become the focal point for professional education in this field.


x
p
c
t
s

Preface

A reader of any reference or text should be concerned about where the book will take him. Hopefully, this preface will answer that question.

Computing has come a long way since the early 1950's when the computer was first introduced on the commercial scene. In the early 1960's, IBM introduced the mass market model 1401 (and later the 1410 and 1440). Both business data processing computer accounting departments and manufacturing departments adopted it. But the applications were different. Accounting was largely the task of processing transactions against a master file of data—time-consuming but relatively straightforward.

Manufacture and production required the storage of the component structure or bill of materials of each product as a data base. Requirements planning, shop routing, and dispatching could be performed from this base. These tasks required a data processing operation that could store and retrieve pieces of data in a disorganized, nonsequential fashion. It was termed *direct access* or *random access processing.*

Since many users were doing approximately the same data processing tasks, the vendors developed software packages as an aid and a marketing inducement. Thus by the time the IBM System 360 series was announced in 1964, these packages had reached a fairly sophisticated level.

These programs served a valuable function by eliminating the necessity of completely replicating programs with each user. But there was a negative effect. The design and development of random access files was slowed considerably. After all, why learn how to do it when all you had to do was learn how to use it.

With the 1970's, computing had reached a crisis. Companies claim they are not receiving a sufficient payoff from their computer investment. There are claims that the computer is idle 50% of the time. There are claims that the manufacturers oversold the product.

To appreciate these relationships, perhaps we should return to a basic understanding of computers. With a thorough grounding in basic concepts, the programmer and systems analyst will be in a position to increase the operating and use efficiency of the key commercial data processing systems and applications of the 1970's.

Therefore, this book on structured information files provides the basics in one aspect of computing. It focuses on how to design a random access file. Yet we do examine two key elements. First, how to get into the file or data base and, second, how to operate within the file to achieve the desired results.

ACKNOWLEDGEMENTS

Many thanks are due to many people. I offer my deepest gratitude to those people who worked with me and aided me at two universities. At the University of Iowa, there were Dr. Peter Schoderbek and Mr. Leon Pearce in the College of Business. Dr. Gerry Weeg and Mr. Paul Wolf in the University of Iowa Computer Center; and at Marquette University Dr. Lloyd Doney, and particularly, Dr. T. R. Martin, Dean of the College of Business Administration. Dr. Martin provided the time, resources, and often critical stimulation to push this work over the peak. The typing of this manuscript was done by Mrs. Gena Lindhorst at Iowa and Mrs. Dolores Rewolinski at Marquette. Without them the war, battle, horse, and shoe would have been lost.

Milwaukee, Wisconsin *Richard A. Kaimann*

List of Tables

Table 1. Randomizing Evaluator Summary
Table 2. Trial 2 Results
Table 3. Trial 7 Results
Table 4. Trial 9 Results
Table 5. Randomizing Evaluator Package Cost Summary
Table 6. Randomizing Evaluator Package Time Summary
Table 7. General Record Format Common Elements
Table 8. Product Data
Table 9. Salesmen
Table 10. Production and Warehouse Facilities
Table 11. Products Sold by Salesmen
Table 12. Warehouses: Quantity On-Hand
Table 13. Facilities: Quantity On-Hand
Table 14. Warehouse Requirements
Table 15. Facilities: Productive Capacity
Table 16. Transportation Costs (as Percent of Produce Cost)
Table 17. Records and Generated Home Addresses
Table 18. Product Record
Table 19. Product Disk Records
Table 20. Salesmen Record
Table 21. Salesmen Disk Records
Table 22. Facility Record
Table 23. Facility Records–Product Data
Table 24. Facility Records–Warehouse Shipping Data
Table 25. Warehouse Record
Table 26. Warehouse Disk Records
Table 27. The Control Record
Table 28. Sales Analyses
Table 29. Stock Status–Quantities On-Hand
Table 30. Transportation Costs: DH-15 Tractor
Table 31. Transportation Costs: DH-17 Cotton Picker
Table 32. Transportation Costs: DH-19 Reaper
Table 33. Transportation Costs: DH-31 Baler
Table 34. Warehouse List Without Structure

Table 35. Warehouse List Alpha Structured
Table 36. Warehouse List Frequency of Occurrence Structured
Table 37. Warehouses Structured by Frequency of Inquiry with Addresses
Table 38. Updated Warehouses Structured by Frequency of Inquiry with Addresses
Table 39. Trial 1 Results
Table 40. Trial 2 Results
Table 41. Trial 3 Results
Table 42. Trial 4 Results
Table 43. Trial 5 Results
Table 44. Trial 6 Results
Table 45. Trial 7 Results
Table 46. Trial 8 Results
Table 47. Trial 9 Results
Table 48. Trial 10 Results

此为试读,需要完整PDF请访问: www.ertongbook.com

Structured Information Files

Contents

One

Introduction 1

A. Purpose of This Book 1
B. The Need for Information 2
C. Hardware Developments 5
D. Unstructured Versus Structured Files 12
E. Summary 15

Two

Background 16

A. The Interdependence of Storage and Retrieval 16
B. Random and Sequential Files 17
C. Characterizing Stored Information 18
D. Software Developments by Vendors 19
E. Software Developments for Specific Applications 22
F. File Organization for Effective Searching 25
G. Entry to the File 28
H. Summary 36

Three

Entry to the File 37

A. The Entry Procedures 37
B. Sample Data for Entry Analysis 38
C. A System of Analysis Programs 38
D. Specific Randomizing Algorithms 41
E. Entry Analysis Results 42
F. Summary 46

Four

Moving About within the File 47

A. The File Design 47
B. Sample Data for the Model 49
C. Developing the Model 53
D. Structured Information File 54
E. Control Records 57
F. Report Generation 59
G. Summary 68

Five

Other Considerations 79

A. Record Placement within the File 79
B. Organizing by Frequency of Occurrence 81
C. File Maintenance 81
D. The Need for a Retrieval Language 84
E. Summary 85

Six

Additional Structured Information File Applications 86

A. The Structured Information File in Educational Administration 86
B. The Structured Information File in Hospital Administration 91
C. An Implication for Processing Large Tape Files 97
D. Summary 102

Seven

Summary 123

Eight

Suggested Problems 126

Bibliography 129

Appendix

The Ten Trial Randomizers and Their Results 135

Index 159

One

Introduction

A. PURPOSE OF THIS BOOK

Information files are a major concern to all corporations. They represent the storehouse of the lifeblood of the organizational body. Without adequate data, the body dies, albeit slowly. With adequate data, the body grows and reaches for maturity and organizational continuity.

These information files reach enormous size. As they grow, the management loses the perspective of what the information was created for, how and where it is kept, and most importantly, what to do with it.

Data processing, with the particular emphasis on computers has added to the complexity. These machines, and the personnel resources that support them, are expensive and often mis-understood. Under the pressures of cost-benefit analyses, the data processing team must produce results.

The key element to producing substantive results is to integrate specific categories of data, judiciously manipulate them, and then feed the management team with meaningful information. The integration of data in a realistic, meaningful situation requires that these data file categories exist. They do exist in the business environment. They also took years and large sums of moeny to develop.

Systems analysts and programmers in the business environment have these data available to them as they develop integrating management information systems. At the same time this is usually the first exposure to computer applications that requires such integration. But, the company has both the data and large scale hardware and storage to handle the task.

Conversely, the student seldom has either the data or large scale data storage capabilities available to him. Therefore, the student is limited to learning just those applications that are either purely computational (such as engineering tasks) or to investigating those information processing techniques where the data are sequential in nature.

The basic purpose of this book is to transcend this learning and conceptual gap. Contrasts will be drawn between typical sequential processing concepts and those where random processing is essential to manipulate a *structured information file.*

A structured information file is a set of categories of data which are related to one another within the category and between categories. These relationships or linkages are constructed to permit logical manipulation in an economic and efficient fashion. The data and their categories are all stored in a random access device.

The method for organizing these data has two very distinct prongs. The first is concerned with entering the file, i.e., how to identify a particular piece of data. The second prong is the movement inside the file. Concepts in the book are presented in a specific context, in terms of actual computer programs. Included are the results of their operation on illustrative applications. The text therefore provides experimental verification of the validity of several of the propositions concerning information files.

The purpose of this book is not an overall view. It is limited to the selection and description of one method and demonstrates that it is feasible. Thus other topics with a global view must be left untreated. The typical user of this book should have had programming exposure—perhaps at this time interested in enlarging his scope of knowledge in the field of data processing by focusing on the area of information files.

Naturally, the audience is not the experienced, high level systems analyst. Indeed, they might be disturbed at some of the omissions. But it should bridge a gap that exists in EDP education between early exposure and practical applications of third generation hardware. Greater degrees of specific information develop as the reader progresses. Those interested in casual study of the subject will do well to read the introductory materials and to follow what others have done in this field. They should not be concerned with the detailed results described in later chapters.

B. THE NEED FOR INFORMATION

Executives and administrators of industry and technology in the last twenty years have faced the problem that with technological progress has come a flood of new information for professionally trained personnel to absorb.[1] Correspondingly, diversity and growth of organizational activity created a mass of administrative data for the manager. Although necessary to translate executive

[1] S. Stuart, "Crystal Balling: New Challenges for Chief Executives," *Infosystems,* January, 1973, p. 22.

duties into action, the vast magnitude of data has surpassed the human processing capability.

Control systems have been established to consolidate data within a firm and to meet the complex needs of management. Multiliferation and elaboration of these systems has been precipitated by human limitations in dealing with great volumes of information. Until the recent past, this problem was usually solved by organizational changes. People and work were organized in such a way that a balance could be set between the data to be processed and the human capabilities to do it.

The application of technology to information processing has been rapidly advanced, both in theory and hardware, since World War II, with the most common manifestation the electronic computer.

Early approaches to the solution of management information needs by computer were simply individual applications converted to automatic data processing. Examples of the piecemeal approach included inventory and material control, accounting and payroll, personnel, sales analysis, etc. More recently, attempts have been made to integrate such fragmented operations into a single computer-based "total information system" for monitoring the activities of the firm.

The total system philosophy has been criticized partly because the approach is inadequate and can lead to additional management problems. The total system approach is not a panacea and one should not assume that system analysis is the most important or only approach. The entire scope of management information requirements cannot be encompassed in one system.[2] The tremendous gap between planning and reality may mean that management is without an information system over a period of years and yet there is no assurance that the system will ever be satisfactory.

In an administrative environment, certain dimensions and features of information requirements cannot be subordinated to the total information system's view. One dimension is the need for varying degrees of data specificity for different functional aspects of management. A more adequate view of management information systems should revolve about a meaningful segregation of these functions. A leading authority has categorized the management functions in three levels, each with significantly different information requirements.

1. Strategic planning which consists of (a) determining corporate policies and objectives, (b) deciding on any changes in these policies and objectives, and (c) deciding on the resources to be devoted to attaining these objectives.

2. Management control, which consists of (a) dividing the strategic plans into logical subdivisions, (b) providing the funds to carry out the subdivisions of

[2] J. Dearden, "MIS is a Mirage," *Harvard Business Review*, January-February, 1972, pp. 90–99.

此为试读,需要完整PDF请访问: www.ertongbook.com