

Edited by Julius T. Tou

Information Systems

COINS

IV

Information Systems

COINS IV

Edited by

Julius T. Tou

*Center for Informatics Research
University of Florida
Gainesville, Florida*

PLENUM PRESS • NEW YORK AND LONDON

Library of Congress Cataloging in Publication Data

Computer and Information Sciences Symposium, 4th, Miami Beach, Fla., 1972.

Information systems.

Includes bibliographies.

1. Electronic digital computers — Programming — Congresses.
2. Management information systems — Congresses.
3. Information storage and retrieval systems — Congresses. I. Tou, Julius T., 1926- ed. II. Title.

QA76.6.C63 1972

029.7

74-4403

ISBN 0-306-35134-X

Proceedings of the Fourth International Symposium on Computer and Information Sciences held in Miami Beach, Florida, December 1972

© 1974 Plenum Press, New York
A Division of Plenum Publishing Corporation
227 West 17th Street, New York, N.Y. 10011

United Kingdom edition published by Plenum Press, London
A Division of Plenum Publishing Company, Ltd.
4a Lower John Street, London W1R 3PD, England

All rights reserved

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the Publisher

Printed in the United States of America

Preface

Ten years ago the first International Symposium on Computer and Information Sciences (COINS-63) was held at Northwestern University. Since that time, computer and information sciences have witnessed a great intensification of research and education. The activities in this field have been significantly broadened and enriched. During this ten-year period, we have organized four COINS symposia to provide a forum for promoting communication among scientists, engineers, and educators in the computer and information science field and to act as a catalyzer for stimulating creative thinking within the community of information processing.

The COINS-72 symposium, which took place in Miami Beach on December 14-16, 1972, under the cosponsorship of the U.S. Army Research Office, the Atomic Energy Commission, and the University of Florida, is the fourth International Symposium on Computer and Information Sciences. The theme of this COINS symposium is *information systems*. This theme has been selected for the following reasons: Information systems have offered widespread applications in education, government, industry, and science. The bulk of research in computer and information science is now geared to the development of improved information systems. A major portion of software engineering is concerned with computer software and sophisticated information system design. It seems logical that a symposium on information systems should follow the preceding software engineering conference.

Last decade, the United States put the space program among its top priorities. The launch of Apollo 17 in December 1972 marked the end of this priority. Now the desire to explore space has given way to other priorities—the betterment of life on earth, the demand for clean air and water, the optimum utilization of energy and natural resources, and the streamlining of municipal governments. Information systems technology holds a key to the solution to these problems. A ten-year space program has generated miles of tapes and volumes of pictures concerning the moon and outer space, and processing them for the discovery of new knowledge and information will keep scientists and engineers busy for the next ten to fifteen years. The Symposium theme has been chosen to reflect the transition from the 60's during which many revolutionary concepts and theories in

computer and information science were developed to the 70's when these concepts will be implemented and extended with even greater accomplishments in solving major national problems.

The COINS-72 program was made up of fifteen technical sessions plus one plenary session and two panel discussion sessions. Among the topics discussed were data base management, software development, information retrieval, pattern recognition, data analysis, urban information systems, and other applications. The authors of these papers and the participants in the panel discussions came from various parts of the United States and from Brazil, Canada, England, France, Germany, India, Israel, Japan, Sweden, and the Soviet Union. The COINS-72 symposium which was attended by delegates from five continents and fifteen countries featured a banquet speech by Dr. Edward E. David, the former Science Advisor to President Nixon, who spoke on *Computing, Information, and International Affairs*.

The papers contained in this book were presented for discussion at the symposium. In order to maintain coherence between the papers and to help the reader locate items of particular interest, the published papers are arranged in logical groups and an index is provided. It is recognized that many other eminent research workers have made significant contributions to the understanding and application of information systems. Unfortunately, the omnipresent tyranny of time and space prohibited the inclusion of their work in the symposium. We sincerely hope that their papers will be presented at the next COINS symposium.

Credit for any success in this symposium must be shared with many people who contributed significantly of their time and talents. In organizing the symposium I, as chairman, received considerable help from the organizing committee, Wayne H. Chen, William F. Kaiser, G. Jack Lipovski, James E. Norman, Milton E. Rose, and James R. Suttle and from the program committee, A. A. Brooks, Ted Codd, George Dodd, Fred Frishman, George Haynam, T. L. Kunii, and C. J. P. Lucena. Much credit is due to our invited reviewers of the symposium papers. My best thanks are also due to Mrs. Grace Searle of the University of Florida and to several members of the Research Technology Division of the Army Research Office at Durham, North Carolina, for their valuable assistance in preparing announcements, programs, and badges, and in arranging and conducting the symposium. To Mrs. Searle, particularly, I owe a debt of gratitude for her able assistance during the time I was bedridden from an automobile accident just prior to the symposium. It was mainly through her efforts that we were able to hold the conference at the scheduled time. It is, however, the authors of the individual papers whose contributions made possible the symposium and the subsequent proceedings. The participation of Vice President Harold P. Hanson of the University of Florida and

Dr. Edward E. David of the White House significantly enhanced the stature of the symposium. To all of them goes the editor's heartfelt thanks and deep appreciation.

Gainesville, Florida

Julius T. Tou

Contents

The Difficulties of Database Management

Gregory C. Everest

1. A Shared Database	1
2. Database Integrity	2
2.1. Facets of Database Integrity	2
2.2. The Means to Database Integrity	10
3. Availability	11
3.1. Diversity of Means	11
3.2. Diversity of Methods	12
3.3. Diversity of Languages	22
3.4. Diversity of Needs	23
4. Evolvability	26
4.1. Changing Technology	27
4.2. Changing User Demands	29
4.3. The Means to Evolvability	30
5. References	32
6. Bibliography	34
6.1. Articles	34
6.2. Books and Other Works	35

Quantized Data Base Systems: A Tutorial

1. Introduction	37
2. The Quantized Model of Data	38
3. A Model of Data Models	40
4. The Hierarchical Approach	43
5. The Network Approach	45
6. A Data Sublanguage for the Relational Model	47
6.1. Relational Algebra	48
6.2. Relational Calculus	50
7. Some Existing Relational Systems	51
8. References	53

Contents

The Objective of Database Management

GORDON C. EVEREST

1. A Shared Database	1
2. Database Integrity	3
2.1. Facets of Database Integrity	5
2.2. The Means to Database Integrity	19
3. Availability	20
3.1. Diversity of Users	20
3.2. Diversity of Modes	22
3.3. Diversity of Languages	23
3.4. Diversity of Needs	25
4. Evolvability	26
4.1. Changing Technology	27
4.2. Changing User Demands	27
4.3. The Means to Evolvability	29
5. References	32
6. Bibliography	34
6.1. Articles	34
6.2. Books and Major Works	35

Relational Data Base Systems: A Tutorial

C. J. DATE

1. Introduction	37
2. The Relational Model of Data	38
3. A Sample Data Model	42
4. The Hierarchical Approach	43
5. The Network Approach	45
6. A Data Sublanguage for the Relational Model	47
6.1. Relational Algebra	48
6.2. Relational Calculus	50
7. Some Existing Relational Systems	51
8. References	53

A Relational Data Management System

V. KEVIN WHITNEY

1. Introduction	55
2. Example	57
3. Application	60
4. Implementation	62
5. Reflections	65
6. References	66

A Data Base Search Problem

FRANK P. PALERMO

1. Introduction	67
1.1. Background	68
1.2. Queries	69
1.3. Assumptions	70
1.4. General Plan	71
1.5. Summary	72
2. Representation of a Query	72
2.1. Introduction	72
2.2. Normalization of β -Expressions	73
2.3. Graphic Representation of a Query	74
2.4. Tabular Representation of a Query	77
2.5. Conclusion	78
3. Improvement of the Reduction Algorithm	78
3.1. Introduction	78
3.2. The Codd Reduction Algorithm	78
3.3. The Evaluation Factors	82
3.4. Improvements on Reduction Algorithm	82
3.5. The Join Algorithm	83
3.6. Improved Reduction Algorithm	84
3.7. Summary	87
4. Algorithm Using Semi-Joins	87
4.1. Introduction	87
4.2. The Semi-Join	87
4.3. The Indirect Join	89
4.4. Target Relations Determined by the T-Table	90
4.5. Exploring a Relation	90
4.6. Estimating Intermediate Storage	90
4.7. The Algorithm Using Semi-Joins	91
4.8. Summary	95

5. Conclusion	96
6. Appendix A. Relational Calculus	96
7. Appendix B. Justification for Reduced Ranges	99
8. References	100

An Experiment with a Relational Data Base System in Environmental Research

KARL SOOP, PER SVENSSON, AND LARS WIKTORIN

1. Introduction	103
1.1. An Environmental Research Problem	103
1.2. Project Background	104
1.3. Problem Characteristics	104
2. Data Processing in an Ecological Research Program	105
2.1. What Activities Are Involved?	105
2.2. Demands on the Software System	106
3. Computer Techniques in the Project	107
3.1. Information Systems Used	107
3.2. Characteristics of IS/1	107
3.3. Some Experiences	110
3.4. An Example	112
4. Conclusion	117
5. References	117

Special Topic Data Base Development

ANDREW J. KASARDA AND DONALD J. HILLMAN

1. Introduction	119
1.1. Content-Induced Partition	120
1.2. Profile-Directed Partition	120
1.3. Data Base Organization	121
2. Content-Induced Partition	121
2.1. Characteristic Weighting Algorithm	122
2.2. Logicostatistical Term Associations	125
2.3. Retrieval Implications	127
3. Profile-Directed Partition	128
3.1. Topic Profile Generation	129
3.2. Term Association Submatrix Partition	131
3.3. Retrieval Implications	132

4. Data Base Organization. Retrieval File Structures	132
5. Summary.....	134
6. References.....	135

BOLTS: A Retrieval Language for Tree-Structured Data Base Systems

W. T. HARDGRAVE

1. Introductory Remarks.....	137
2. Preliminary Definitions.....	138
3. Retrieval Procedure.....	139
4. Examples of Retrievals in SET-BARS and TREE-BARS.....	140
4.1. An Example of the Set-Theoretic System.....	140
4.2. An Example of the Tree-Theoretic System.....	143
5. Definition of BOLTS.....	145
5.1. Set Manipulation Functions.....	145
5.2. Node Extraction Functions.....	146
5.3. Selection and Qualification in BOLTS.....	146
5.4. Examples of SELECT, ADJUST, QUALIFY, and TYPE.....	147
6. Tree Operations in BOLTS.....	147
6.1. Preliminary Theorems.....	147
6.2. Tree Intersection in BOLTS.....	148
6.3. Tree Complement in BOLTS.....	151
6.4. Examples of Tree Operations in BOLTS.....	151
7. The "HAS Clause" in BOLTS.....	152
7.1. An Example of Sibling Retrieval.....	153
7.2. An Example of Indirect Ancestor Retrieval.....	154
7.3. An Additional Capability in BOLTS.....	154
8. Concluding Remarks.....	155
9. References.....	158

An Algorithm for Maintaining Dynamic AVL Trees

JAMES R. VAN DOREN AND JOSEPH L. GRAY

1. Introduction.....	161
2. AVL Trees.....	162
3. Searching.....	163
4. Insertion.....	166
5. Deletion.....	171
6. The Implemented Algorithm.....	177
7. Comparison with Binary Search Trees of Bounded Balance.....	178
8. References.....	180

SPIRAL's Autoindexing and Searching Algorithms

LESLIE E. WEST

1. Introduction	181
2. Indexing and Storage System	182
2.1. Exclusion Words	183
2.2. Suffix Truncation	184
2.3. Encoding for Vocabulary Indices	185
2.4. Encoding for Word Usage Patterns	186
3. Inquiry Form	186
4. Inquiry Compilation	189
5. Retrieval Process	190
5.1. Type 1 Processing	191
5.2. Type 3 Processing	191
5.3. Type 5 Processing	192
5.4. Type 7 Processing	194
6. Conclusion	195
7. References	196

SEFIRE: A Sequential Feedback Interactive Retrieval System

JULIUS T. TOU AND FRED R. SUTTON

1. Introduction	197
2. Characteristics of Interactive Information Retrieval System	199
3. Hierarchical Category Files	200
4. Software Design	202
4.1. Design Principles	202
4.2. System Tables	207
5. Experimental Results	213
6. Conclusions	216
7. References	216

An Analysis of Document Retrieval Systems Using a Generalized Model

CAROLYN J. CROUCH AND DONALD B. CROUCH

1. Introduction	219
2. The Generalized Model	220
2.1. User	221
2.2. Logical Processor	222

2.3. Selector	222
2.4. Descriptor File	223
2.5. Locator	223
2.6. Document File	223
2.7. Data	224
2.8. Analysis	224
3. Analysis of Implemented Systems	225
3.1. Query System	225
3.2. GIPSY	227
3.3. BIRS	230
3.4. SMART	232
4. Summary	236
5. References	236

Information Systems for Urban Problem Solvers

MANFRED KOCHEN

1. Introduction: Recognition of a Need for Urban Information Systems	239
2. A Typology of Problems: Information Systems for Problem-Solving	241
3. Information Systems for Well-Defined Problems	246
4. Functions of an Information System for Ill-Structured Problems	249
5. Design Principles	252
6. Conclusions and Recommendations	255
7. Appendix A: A Model for the Simplest Shopping Problem	256
8. Appendix B: Consequences of a Decision by People Who Have Undesirable Genes Not to Have Offspring	258
9. References	260

EMISARI: A Management Information System Designed to Aid and Involve People

R. L. RENNER, R. M. BECHTOLD, C. W. CLARK,
N. H. GOLDSTEIN, D. O. MARBRAY, AND R. L. WYNN

1. Introduction	263
2. Description of System	266
2.1. User's Guide, Description, and Explanation Choices	266
2.2. Agencies and Contacts	270

2.3. Messages and Communication.....	271
2.4. Estimates, Programs, and Tables.....	272
2.5. Text Files.....	275
2.6. Special Features.....	278
3. Role of the Monitor.....	280
4. Implementation Features.....	283
4.1. Use of XBASIC.....	283
4.2. Files and Adaptive Index.....	286
4.3. Data Survivability.....	287
5. References.....	289

Transferability and Translation of Programs and Data

E. H. SIBLEY AND A. G. MERTEN

1. Introduction.....	291
2. Aspects of Language Translation.....	293
3. Aspects of Data Translation.....	295
3.1. Definitions of Data Terms.....	296
3.2. A Model of Data Accessing.....	297
3.3. Generalized Data Access and Translation.....	299
4. Interdependence of Program and Data Translation.....	302
5. Features of Program and Data Translation.....	304
5.1. Logical Elements of a Program Translator.....	304
5.2. Logical Elements of a Data Translator.....	307
5.3. Uniqueness of Translation.....	307
6. Conclusions.....	307
7. References.....	309

Processing Systems Optimization through Automatic Design and Reorganization of Program Modules

J. F. NUNAMAKER, JR., W. C. NYLIN, JR., AND B. KONSYNSKI, JR.

1. Introduction.....	311
2. Methodology.....	312
3. Definitions.....	313
4. Process Grouping Concept.....	319
5. Process Grouping Determination.....	322
5.1. Generation of Feasible Process Groupings to Form Modules.....	324

5.2. Generation of Alternative System Designs.....	327
5.3. Transport Volume Savings Calculation.....	329
6. Combining Processes.....	329
7. Example.....	331
8. Conclusions.....	335
9. References.....	336

Verification and Checking of APL Programs

SUSAN GERHART

1. Introduction.....	337
2. Proving Assertions about APL Programs.....	340
3. Verification of Constraints of APL Programs.....	341
3.1. Straight-Line Programs with Assertions.....	343
3.2. Programs with Branches and Assertions.....	344
3.3. Programs with Branches and No Assertions.....	344
4. Summary and Conclusions.....	345
5. References.....	346

G/PL/I: Extending PL/I for Graph Processing

C. SARAIVA DOS SANTOS AND A. L. FURTADO

1. Introduction.....	347
2. An Informal Description of the Extension.....	348
3. Implementation Considerations.....	353
4. An Example.....	354
5. Directions for Further Developments.....	355
6. Appendix.....	356
7. References.....	359

A Unified Approach to the Evaluation of a Class of Replacement Algorithms

EROL GELENBE

1. Introduction.....	361
2. Definition of Basic Concepts.....	363
3. Random, Partially Preloaded Algorithms.....	367

4. Proof of Theorem 2.....	368
5. The Algorithms RAND and FIFO.....	371
6. Appendix. Proof of Lemma 1.....	373
7. References.....	375

Quantitative Timing Analysis and Verification for File Organization Modeling

V. Y. LUM, M. E. SENKO, H. LING, AND J. H. BARLOW

1. Introduction.....	377
2. General Description and Organization of the Model.....	378
3. Techniques of Analysis.....	379
4. Experimental Evaluation of the Timing Equations.....	383
5. Conclusion.....	385
6. References.....	386

A Mathematical Model for Computer-Assisted Document Creation

SHMUEL S. OREN

1. Introduction.....	387
2. Description of the Model and Its Mathematical Representation.....	388
3. Optimal Operation.....	392
4. A Special Case: "Ideal Operator—Exponential File".....	394
5. Application to System Design.....	396
6. Conclusions.....	399
7. Appendix.....	399
8. References.....	401

Representing Geographic Information for Efficient Computer Search

R. D. MERRILL AND M. TANNENBAUM

1. Introduction.....	403
1.1. Subject.....	403
1.2. Examples.....	404
2. Representation Technique.....	404
2.1. Basic Data Structure.....	404

2.2. Properties of the TCB Structure.....	407
2.3. Representing Regional Information.....	410
2.4. Representing Contour Map Information.....	412
3. Retrieval Applications.....	415
3.1. Geographic Information System.....	415
3.2. Terrain Coverage Information for Microwave Radiometer Image Prediction Model.....	417
3.3. Terrain Relief Information for Radar Image Prediction Model.....	419
4. Summary.....	421
5. Appendix. Contour Map Search List Determination.....	422
6. References.....	423

A Syntactic Pattern Recognition System with Learning Capability

H. C. LEE AND K. S. FU

1. Introduction.....	425
2. Design Concepts and Overall System Description.....	428
3. Learning of Pattern Grammar.....	429
4. Learning of Production Probabilities.....	433
5. Computational Results.....	437
6. Conclusion.....	448
7. References.....	448

Optimization in Nonhierarchic Clustering

EDWIN DIDAY

1. Introduction.....	451
1.1. The Problem.....	451
1.2. The Dynamic Clusters Method.....	452
1.3. Synthetic Study of the Solutions Obtained.....	453
2. Some Notations and Definitions.....	453
3. Constructing the Triplets (f, g, W)	454
3.1. General Formulation.....	454
3.2. The Different Variants and a Comparison of Some of Interest.....	455
3.3. Construction of Triplets That Make the Sequence u_n Decreasing.....	457

4. The Structure of $\mathbb{L}_k$, $\mathbb{P}_k$, V_k and Optimality Properties.....	458
4.1. The Nonbiased Elements.....	458
4.2. The Impasse Elements.....	458
5. Searching for Invariants.....	460
5.1. Measure of the Rooted Trees.....	460
5.2. Strong Forms, Fuzzy Sets, and Information.....	461
5.3. Global Optimum of V_k	463
5.4. Approaching the Global Optimum by Changing Trees..	463
6. Programming the Tables of the Strong Forms and the Heuristic Interpretation.....	464
7. Examples of Applications.....	465
7.1. The Artificial Example of Ruspini.....	465
7.2. Classifying the Soundings of a Mine for Its Minerals....	471
7.3. Study of Serum Protein Disturbance in Clinical Pathology	473
8. Conclusion.....	474
9. Appendix A.....	474
10. Appendix B.....	476
11. Appendix C.....	477
12. References.....	478

Nonparametric Learning Using Contextual Information

MASAMICHI SHIMURA,

1. Introduction.....	481
2. Structure of the Machine.....	482
3. Nonparametric Learning.....	484
4. Computer Simulation.....	489
5. Concluding Remarks.....	492
6. References.....	493