Contents

1 Communicating NASA's Science to the Public
by France Cordova 1
1.1 Introduction 1

2 The Library of Congress's National Digital Library Program
by James H. Billington 17
2.1 Introduction 17

3 The Universal Library: Intelligent Agents and Information on Demand
by Raj Reddy 27
3.1 Introduction 27
3.2 A Future Scenario 27
3.3 Conventional vs. Digital Libraries 29
3.4 The Economic Model 30
3.5 The Content and Infrastructure Model 31
3.6 The Home Information System Model 32
3.7 The Operating System Model 32
3.8 The Human Computer Interaction Model 33
3.9 Conclusion 33

4 Building a Scalable America
by Larry Smarr 35
4.1 Introduction 35

I Visualization

5 Video as Scholarly Material in the Digital Library
by Wayne Wolf, Bede Liu, Andrew Wolfe, MinervaYeung, Boon-Lock Yeo, and Daniel Markham 45
5.1 Moving images as scholarly material 45
5.2 Digital Video Libraries 46
5.3 Methodologies and algorithms for digital video libraries 46
5.4 Conclusions 53

6 Digital Libraries for Electronic News
by Michael A. Shepherd, C.R. Watters, and F.J. Burkowski 55
6.1 Introduction 55
6.2 Newspaper Databases 57
6.3 Electronic News 57
6.4 Features of an Electronic News System 58
6.5 Electronic News as a Digital Library 58
6.5.1 Attributes of News Items 58
6.5.2 Integration of News Delivery and a Digital Library 59
6.6 Summary 62

7 WebJournal: Visualization of a Web Journey
by Bipin C. Desai and Stan Swiercz 63
7.1 Introduction 63
7.2 Journaling a Web Journey 64
7.2.1 Layout and Semantics of WebJournal 64
7.2.2 Semantics of Nodes 67
7.2.3 Semantics of Edges 67
7.2.4 Handling Multiple Browser Windows 69
7.3 Managing the WebJournal Real Estate 69
7.4 Marking Nodes, Saving, Reloading and Printing 72
7.4.1 Exploding a Node 73
7.4.2 Marked Nodes and Printing 75
7.5 WebJournal Internals, communications. 79
7.6 Conclusions 79
7.7 Acknowledgment 80

II Document Handling and Information Retrieval

8 Uniform Structured Document Handling using a Constraint-based Object Approach
by Anisoara Nica, Elke Angelika Rundensteiner 83
8.1 Introduction 84
8.1.1 Structured Document Management 84
8.1.2 The Constraint-based Approach 85
8.1.3 Overview 86
8.2 Project Background 86
8.3 The Constraint-Based Object Model 88
8.4 Mapping SGML Model Groups into the Constraint Based Model 90
8.5 Schema Merging Using the Constraint Model 97
8.6 Related Work 99
8.7 Conclusions 100

9 Digital Software and Data Repositories for Support of Scientific Computing
by Ronald Boisvert, Shirley Browne, Jack Dongarra, and Eric Grosse **103**
9.1 Introduction . . . 103
9.2 Characteristics of Some Existing Software Repositories . . . 104
9.2.1 Netlib . . . 104
9.2.2 The National HPCC Software Exchange (NHSE) . . . 105
9.2.3 GAMS Virtual Repository . . . 106
9.3 Indexing and Searching of Software Objects . . . 107
9.3.1 Data Models . . . 107
9.3.2 Software Classification . . . 107
9.3.3 Search Interfaces . . . 108
9.4 Retrieval of Software Objects . . . 109
9.4.1 Downloading Files . . . 109
9.4.2 Templates and Archetypes . . . 110
9.4.3 Remote Execution . . . 111
9.4.4 Change Notification . . . 111
9.5 Access to Scientific Data . . . 112
9.6 Integration with Document Digital Libraries . . . 113

10 Semantic Hypermedia Retrieval in Digital Libraries
by Stephan Wiesener, Wolfgang Kowarschick, Pavel Vogel, and Rudolf Bayer **115**
10.1 Introduction . . . 115
10.2 The OMNIS Digital Library System . . . 118
10.2.1 OMNIS System Architecture . . . 118
10.2.2 Archiving and Retrieval . . . 120
10.2.3 OMNIS in Practice . . . 121
10.3 Towards Knowledge-based Navigation in Digital Libraries . . . 121
10.3.1 System Architecture . . . 123
10.3.2 Documents and Semantic Knowledge Bases . . . 125
10.3.3 Queries and Navigation . . . 127
10.4 Summary and Outlook . . . 128

11 Fuzzy Full-Text Searches in OCR Databases
by Andreas Myka and Ulrich Güntzer **131**
11.1 Introduction . . . 131
11.2 Search mechanisms . . . 133
11.2.1 Exact match . . . 133
11.2.2 Canonical forms . . . 134
11.2.3 N-grams . . . 137
11.2.4 Linear scanning . . . 137
11.3 Experiments . . . 140
11.3.1 Environment . . . 140

11.3.2 Methodology 140
11.3.3 Results 141
11.4 Summary and outlook 145

III Network-Based Information and Resource Discovery

12 Data Discovery in Large Scale Heterogeneous and Autonomous Databases
by Athman Bouguettaya and Stephen Milliner **149**
12.1 Background 149
12.1.1 Motivation 149
12.2 Related Work 151
12.3 The FINDIT Database Discovery System 153
12.3.1 The FINDIT Architecture Level 155
12.3.2 The FINDIT Interaction/Negotiation Level 157
12.3.3 The FINDIT Interoperability Level 159
12.3.4 The FINDIT Exploration Level 160
12.4 Implementation Overview 163
12.5 Discussion and Future Directions 166

13 An Intelligent Agent for the K-12 Educational Community
by Mark E. Rorvig, Mark W. Hutchison, Robert O. Shelton, Stephanie L. Smith and Marwan E. Yazbeck **167**
13.1 Introduction 167
13.2 The Agent 168
13.3 Acknowledgements 176
13.4 Appendix A - Example of Question and Answer Responses from November and April 176

14 Interface Issues for Interactive Multimedia Documents
by Robert B. Allen **179**
14.1 Introduction 179
14.1.1 Multimedia Documents 179
14.1.2 Multimedia Browsers 180
14.1.3 Internal Representations 180
14.2 Presentation Issues For Multimedia Documents 180
14.2.1 Tables of Contents (TOCs) 181
14.2.2 Linking 181
14.2.3 Searching 181
14.2.4 Supporting Partially Guided Tours 182
14.2.5 Locking Concurrent Multimedia Streams 182
14.3 TOC Interface Implementations 183
14.3.1 Corpora 183
14.3.2 Timeline-based Audio-Slideshow Browser 184

14.3.3 TOC-based View-graph and Audio Browser 185
14.3.4 TOC-based Video Browser 185
14.3.5 Visual TOC Browser 186
14.4 Discussion 187
14.4.1 Other Media, Other Structures, and Other Widgets . . . 187
14.4.2 Possible Social Impact of Multimedia Lectures 188
14.4.3 Envoi 188
14.5 Acknowledgements 189

15 Searching and Discovery of Resources in Digital Libraries by Nahum Gershon, William Ruh, Joshua LeVasseur, Joel Winstead, and Adrienne Kleiboemer 191
15.1 Introduction 191
15.2 Browsing Through Hyperspace Without Being Lost 193
15.3 Overcoming the Rigidity of the WWW– Building One Owns Information Hyperspace 193
15.4 Finding New Information in Retrieved Documents– Aggregating Relevant Information 196
15.5 Classification of Information and Metadata Extraction 197
15.6 In Conclusion 198
15.7 Acknowledgments 198

IV Design Issues and Prototyping

16 The Almaden Distributed Digital Library System by David M. Choy, Richard Dievendorff, Cynthia Dwork, Jeffrey B. Lotspiech, Robert J. T. Morris, Norman J. Pass, Laura C. Anderson, Alan E. Bell, Stephen K. Boyer, Thomas D. Griffin, Bruce A. Hoenig, James M. McCrossin, Alex M. Miller, Florian Pestoni and Deidra S. Picciano 203
16.1 Introduction 203
16.2 The Library Principals and their Needs 207
16.2.1 The consumer 207
16.2.2 The librarian 207
16.2.3 The primary publisher 208
16.2.4 The secondary publisher 208
16.3 Architecture of the Almaden Distributed Digital Library System 209
16.3.1 The Source Server 212
16.3.2 The Customer Server 213
16.3.3 The Client 215
16.4 Security and Rights Management 215
16.5 Summary and Future Work 218

17 Alexandria Digital Library: Rapid Prototype and Metadata Schema
by Christoph Fischer, James Frew, Mary Larsgaard, Terence R. Smith and Qi Zheng 221
17.1 Introduction . . . 221
17.2 Goals, Strategy, and General Architecture of ADL . . . 222
17.2.1 The Strategy and General Architecture of ADL . . . 223
17.3 The Rapid Prototype System for the ADL . . . 225
17.3.1 Classes of Queries and the Functional Architecture of the RPS . . . 226
17.4 The Interfaces to the ADL Rapid Prototype . . . 227
17.4.1 The User Interface to the RPS . . . 228
17.5 The Catalog and Metadata for the ADL Rapid Prototype . . . 231
17.5.1 A Comparison of the USMARC and FGDC Metadata standards . . . 231
17.5.2 General Issues in Implementing USMARC and FGDC Standards . . . 233
17.5.3 Combining the FGDC and USMARC Standards in the RPS 234
17.6 The Ingest and Storage Components of the Rapid Prototype . . 237
17.7 Software and Hardware Components of the ADL Rapid Prototype 238
17.8 Unique Features of the RPS, Lessons Learned, and Next Steps . 239
17.8.1 The Next Stage in constructing ADL . . . 239
17.9 Summary . . . 240

18 The ELINOR Electronic Library
by Dian G. Zhao and A. Ramsden 243
18.1 Introduction . . . 243
18.2 Using ELINOR . . . 245
18.2.1 The Database . . . 245
18.2.2 Browsing and Reading Documents . . . 246
18.2.3 Searching for Documents . . . 246
18.2.4 Printing Documents . . . 249
18.3 User Study . . . 250
18.4 The Systems Aspect of ELINOR . . . 250
18.4.1 The System Architecture . . . 250
18.4.2 Usage Statistics Collection . . . 252
18.4.3 Usage Statistics Management . . . 252
18.4.4 Usage Statistics Reporting . . . 255

19 Dienst: Building a Production Technical Report Server
by James R. Davis, Carl Lagoze, and Dean B. Krafft 259
19.1 Introduction . . . 259
19.2 Overview of Dienst . . . 259
19.3 Experience Gained . . . 260
19.3.1 Dienst and the World Wide Web . . . 260
19.3.2 Copyright Issues . . . 262

19.3.3 Document Submission and Management 264
19.3.4 Providing Documents in Multiple Formats 265
19.4 Ongoing Issues 266
19.4.1 Interoperability Among Heterogeneous Search Engines . . 266
19.4.2 Heterogeneous Servers and the Dynamic User Interface . 267
19.4.3 Reliability 267
19.4.4 Document Structure 268
19.4.5 Technology Transfer 268
19.5 Future Plans 269
19.6 Acknowledgements 270

Bibliography **273**