

- List of figures p. xiii
- List of tables p. xix
- Foreword Edward A. Fox p. xxi
- Preface p. xxv
- 1. Orientation: The world of digital libraries p. 1
- Example 1 Supporting human development p. 1
- Example 2 Pushing on the frontiers of science p. 2
- Example 3 Preserving a traditional culture p. 3
- Example 4 Exploring popular music p. 4
- The scope of digital libraries p. 5
- 1.1 Libraries and digital libraries p. 5
- 1.2 The changing face of libraries p. 8
- In the beginning p. 10
- The information explosion p. 11
- The Alexandrian principle p. 14
- Early technodreams p. 15
- The library catalog p. 16
- The changing nature of books p. 17
- 1.3 Digital libraries in developing countries p. 20
- Disseminating humanitarian information p. 21
- Disaster relief p. 21
- Preserving indigenous culture p. 22
- Locally produced information p. 22
- The technological infrastructure p. 23
- 1.4 The Greenstone software p. 24
- 1.5 The pen is mighty: Wield it wisely p. 28
- Copyright p. 29
- Collecting from the Web p. 31
- Illegal and harmful material p. 34
- Cultural sensitivity p. 34
- 1.6 Notes and sources p. 35
- 2. Preliminaries: Sorting out the ingredients p. 39
- 2.1 Sources of material p. 40
- Ideology p. 41
- Converting an existing library p. 42
- Building a new collection p. 43
- Virtual libraries p. 44
- 2.2 Bibliographic organization p. 46
- Objectives of a bibliographic system p. 47
- Bibliographic entities p. 48
- 2.3 Modes of access p. 55
- 2.4 Digitizing documents p. 58
- Scanning p. 59
- Optical character recognition p. 61
- Interactive OCR p. 62
- Page handling p. 67

- Planning an image digitization project p. 68
- Inside an OCR shop p. 69
- An example project p. 70
- 2.5 Notes and sources p. 73
- 3. Presentation: User interfaces p. 77
- 3.1 Presenting documents p. 81
  - Hierarchically structured documents p. 81
  - Plain, unstructured text documents p. 83
  - Page images p. 86
  - Page images and extracted text p. 88
  - Audio and photographic images p. 89
  - Video p. 91
  - Music p. 92
  - Foreign languages p. 93
- 3.2 Presenting metadata p. 96
- 3.3 Searching p. 99
  - Types of query p. 100
  - Case-folding and stemming p. 104
  - Phrase searching p. 106
  - Different query interfaces p. 108
- 3.4 Browsing p. 112
  - Browsing alphabetical lists p. 113
  - Ordering lists of words in Chinese p. 114
  - Browsing by date p. 116
  - Hierarchical classification structures p. 116
- 3.5 Phrase browsing p. 119
  - A phrase browsing interface p. 119
  - Key phrases p. 122
- 3.6 Browsing using extracted metadata p. 124
  - Acronyms p. 125
  - Language identification p. 126
- 3.7 Notes and sources p. 126
  - Collections p. 126
  - Metadata p. 127
  - Searching p. 127
  - Browsing p. 128
- 4. Documents: The raw material p. 131
- 4.1 Representing characters p. 134
  - Unicode p. 137
  - The Unicode character set p. 138
  - Composite and combining characters p. 143
  - Unicode character encodings p. 146
  - Hindi and related scripts p. 149
  - Using Unicode in a digital library p. 154
- 4.2 Representing documents p. 155
  - Plain text p. 156

- Indexing p. 157
- Word segmentation p. 160
- 4.3 Page description languages: PostScript and PDF p. 163
- PostScript p. 164
- Fonts p. 170
- Text extraction p. 173
- Using PostScript in a digital library p. 178
- Portable Document Format: PDF p. 179
- PDF and PostScript p. 183
- 4.4 Word-processor documents p. 184
- Rich Text Format p. 185
- Native Word formats p. 191
- LaTeX format p. 191
- 4.5 Representing images p. 194
- Lossless image compression: GIF and PNG p. 195
- Lossy image compression: JPEG p. 197
- Progressive refinement p. 203
- 4.6 Representing audio and video p. 206
- Multimedia compression: MPEG p. 207
- MPEG video p. 210
- MPEG audio p. 211
- Mixing media p. 212
- Other multimedia formats p. 214
- Using multimedia in a digital library p. 215
- 4.7 Notes and sources p. 216
- 5. Markup and metadata: Elements of organization p. 221
- 5.1 Hypertext markup language: HTML p. 224
- Basic HTML p. 225
- Using HTML in a digital library p. 228
- 5.2 Extensible markup language: XML p. 229
- Development of markup and stylesheet languages p. 230
- The XML metalanguage p. 232
- Parsing XML p. 235
- Using XML in a digital library p. 236
- 5.3 Presenting marked-up documents p. 237
- Cascading style sheets: CSS p. 237
- Extensible stylesheet language: XSL p. 245
- 5.4 Bibliographic metadata p. 253
- MARC p. 254
- Dublin Core p. 257
- BibTeX p. 258
- Refer p. 260
- 5.5 Metadata for images and multimedia p. 261
- Image metadata: TIFF p. 262
- Multimedia metadata: MPEG-7 p. 263
- 5.6 Extracting metadata p. 266

- Extracting document metadata p. 267
- Generic entity extraction p. 268
- Bibliographic references p. 270
- Language identification p. 270
- Acronym extraction p. 271
- Key-phrase extraction p. 273
- Phrase hierarchies p. 277
- 5.7 Notes and sources p. 280
- 6. Construction: Building collections with Greenstone p. 283
- 6.1 Why Greenstone? p. 285
  - What it does p. 285
  - How to use it p. 288
- 6.2 Using the Collector p. 292
  - Creating a new collection p. 293
  - Working with existing collections p. 300
  - Document formats p. 301
- 6.3 Building collections manually: A walkthrough p. 302
  - Getting started p. 303
  - Making a framework for the collection p. 304
  - Importing the documents p. 305
  - Building the indexes p. 307
  - Installing the collection p. 308
- 6.4 Importing and building p. 309
  - Files and directories p. 310
  - Object identifiers p. 312
  - Plug-ins p. 313
  - The import process p. 314
  - The build process p. 317
- 6.5 Greenstone archive documents p. 319
  - Document metadata p. 320
  - Inside the documents p. 322
- 6.6 Collection configuration file p. 323
  - Default configuration file p. 324
  - Subcollections and supercollections p. 325
- 6.7 Getting the most out of your documents p. 327
  - Plug-ins p. 327
  - Classifiers p. 336
  - Format statements p. 342
- 6.8 Building collections graphically p. 349
- 6.9 Notes and sources p. 353
- 7. Delivery: How Greenstone works p. 355
- 7.1 Processes and protocols p. 356
  - Processes p. 357
  - The null protocol implementation p. 357
  - The Corba protocol implementation p. 359
- 7.2 Preliminaries p. 360

- The macro language p. 360
- The collection information database p. 369
- 7.3 Responding to user requests p. 372
  - Performing a search p. 375
  - Retrieving a document p. 376
  - Browsing a hierarchical classifier p. 377
  - Generating the home page p. 378
  - Using the protocol p. 378
  - Actions p. 384
- 7.4 Operational aspects p. 385
  - Configuring the receptionist p. 386
  - Configuring the site p. 391
- 7.5 Notes and sources p. 392
- 8. Interoperability: Standards and protocols p. 393
  - 8.1 More markup p. 395
    - Names p. 395
    - Links p. 397
    - Types p. 402
  - 8.2 Resource description p. 408
    - Collection-level metadata p. 410
  - 8.3 Document exchange p. 413
    - Open eBook p. 414
  - 8.4 Query languages p. 419
    - Common command language p. 419
    - XML Query p. 422
  - 8.5 Protocols p. 426
    - Z39.50 p. 427
      - Supporting the Z39.50 protocol p. 429
      - The Open Archives Initiative p. 430
      - Supporting the OAI protocol p. 433
  - 8.6 Research protocols p. 434
    - Dienst p. 435
      - Simple digital library interoperability protocol p. 436
      - Translating between protocols p. 437
      - Discussion p. 438
  - 8.7 Notes and sources p. 440
- 9. Visions: Future, past, and present p. 443
  - 9.1 Libraries of the future p. 445
    - Today's visions p. 445
    - Tomorrow's visions p. 448
    - Working inside the digital library p. 451
  - 9.2 Preserving the past p. 454
    - The problem of preservation p. 455
    - A tale of preservation in the digital era p. 456
    - The digital dark ages p. 457
    - Preservation strategies p. 459

- 9.3 Generalized documents: A challenge for the present p. 462
- Digital libraries of music p. 462
- Other media p. 466
- Generalized documents in Greenstone p. 469
- Digital libraries for oral cultures p. 471
- 9.4 Notes and sources p. 474
- Appendix Installing and operating Greenstone p. 477
- Glossary p. 481
- References p. 489
- Index p. 499
- About the authors p. 517