CONTENTS

Preface to the Second Edition ..... *xi*
Foreword from the First Edition ..... *xvii*
Preface from the First Edition ..... *xxi*
List of Figures ..... *xxxiii*
List of Tables ..... *xxxix*

Chapter 1 Biomedical Data *1*
1.1 The Nature and Representation of Biomedical Data 5
1.1.1 What Can Be Represented in a Computer? 18
1.1.2 Cells, DNA, Proteins, and the Genetic Code 37
1.1.3 Anatomy 54
1.1.4 Medical Laboratory Data 56
1.1.5 Medical Images 63
1.1.6 Metadata 83
1.2 Objects, Metadata, and Serialization 87
1.2.1 A Simple Solution Using Tags (Keywords) 88
1.2.2 An Object-Oriented Design 90
1.2.3 A Better Solution Using Metaobjects 102
1.2.4 Medical Images: Incorporating Binary Data 111
1.2.5 Drug Interactions: Organizing Drugs into Classes 118
1.2.6 XML 124
1.3 Database Concepts and Systems 133
1.3.1 The Relational Model 134
1.3.2 Indexing and Tree Structures 143
1.3.3 Time-Oriented Data and Interval Trees 150
1.3.4 The Entity-Attribute-Value (EAV) Model 155
1.3.5 Online Analytical Processing (OLAP) 156
1.4 Data Quality 165
1.5 Data, Information, and Knowledge 169
1.6 Summary 173
1.7 Exercises 173

Chapter 2 Symbolic Biomedical Knowledge *179*
2.1 Biomedical Theories and Computer Programs 182
2.1.1 A World Class Reasoning Example 182

2.1.2 Biological Examples 183
2.1.3 Symbolic Theories 184
2.2 Logic and Inference Systems 187
2.2.1 Predicate Calculus 189
2.2.2 Unsound Inference: Abduction and Induction 218
2.2.3 First Order Logic (FOL) 219
2.2.4 Rule-Based Programming for FOL 230
2.2.5 Example: Predicting Drug Interactions by Logical Inference 237
2.2.6 Truth Maintenance Systems 242
2.2.7 Electronic Medical Records and Guidelines 254
2.2.8 Example: Cancer Radiotherapy Planning 255
2.2.9 Example: Rule-Based Nursing Knowledge 272
2.2.10 Limitations of FOL 277
2.3 Frames, Semantic Nets, and Ontologies 281
2.3.1 A Simple Frame System 281
2.3.2 Extensions to Frame Systems 303
2.3.3 Frame System Implementations, Interfaces, and APIs 306
2.3.4 The UW Foundational Model of Anatomy 307
2.3.5 Representing Drug Information in Frames 313
2.3.6 Frames and Object-Oriented Programming Languages 316
2.3.7 Example: The Gene Ontology 318
2.4 Description Logics 321
2.4.1 Description Logics and Notation Systems 325
2.4.2 Example: GALEN 326
2.4.3 The Semantic Web 330
2.4.4 Example: Enhancing Queries With LOINC 333
2.4.5 Back to the Future 335
2.5 Summary 335
2.6 Exercises 337

Chapter 3 Probabilistic Biomedical Knowledge 343

3.1 Probability and Statistics 344
3.1.1 Probability 345
3.1.2 Statistics 346
3.1.3 The Laws of Probability 347
3.1.4 Conditional Probability 350
3.1.5 Independence 351
3.1.6 Random Variables and Estimation 352
3.2 Bayes' Rule, Bayes Nets, and Probabilistic Inference 364
3.2.1 Simple Bayesian Inference 365
3.2.2 Non-Boolean Variables 367
3.2.3 Bayes Nets 368
3.2.4 Example: Radiation Therapy Planning 373
3.3 Utility and Decision Modeling 375
3.3.1 A Decision Analysis Vignette 376

3.3.2 Graph Structure and Probability Assignment 377
3.3.3 Determining Utilities 378
3.3.4 Computation of Expected Values 381
3.4 Information Theory 382
3.4.1 Encoding of Messages 382
3.4.2 Entropy and Information 386
3.4.3 Efficient Encoding 388
3.4.4 Error Detection and Correction 391
3.4.5 Information Theory in Biology and Medicine 391
3.5 Summary 395
3.6 Exercises 395

Chapter 4 Biomedical Information Access 397
4.1 Information Retrieval Systems 398
4.1.1 IR System Design 402
4.1.2 Indexing 403
4.1.3 Processing Queries 410
4.1.4 Searching and Matching 410
4.1.5 Ranking the Results 410
4.1.6 Performance Evaluation 411
4.2 Beyond Text Documents 413
4.2.1 Biomedical Images 413
4.2.2 Electronic Medical Records 414
4.3 Networks and Biomedical Data Communication 415
4.3.1 Networks and Protocols 418
4.3.2 HL7: Clinical Data Encoding and Transmission 424
4.3.3 DICOM: Medical Image Data Encoding and Transmission 433
4.3.4 Network Access to Biomedical Documents and Data 445
4.3.5 Network Access to an Ontology: The FMA 458
4.4 Summary 476
4.5 Exercises 476

Chapter 5 Computational Models and Methods 479
5.1 Computing with Genes, Proteins, and Cells 480
5.1.1 Mapping Genes to Proteins 480
5.1.2 Computing with Protein Sequences 485
5.2 Search 496
5.2.1 A General Framework for Search 497
5.2.2 Uninformed Search Strategies 503
5.2.3 Informed Search Strategies 505
5.2.4 Path Search 508
5.2.5 Example: DNA Sequence Alignment 509
5.2.6 Example: Biochemical Pathways 515
5.3 Natural Language Processing 523
5.4 State Machines and Dynamic Models 525

5.4.1 State Machines 526
5.4.2 A Simple State Machine 529
5.4.3 Example: Simulating the Cell Cycle in Yeast 533
5.4.4 Example: Implementing DICOM with a State Machine 540
5.5 Stochastic Processes 562
5.5.1 Markov Chains 563
5.5.2 Example: Prostate Cancer 565
5.5.3 Example: Radiotherapy Clinical Target Volume 567
5.6 Summary 575
5.7 Exercises 576

Chapter 6 Biomedical Data, Knowledge, and Systems in Context 579
6.1 A Bit of Personal History and Perspective 580
6.2 Medical Device Safety 583
6.2.1 Radiation Therapy Planning Systems 583
6.2.2 Who is Master, the Human or the Computer? 585
6.2.3 The Therac-25: A Real-Time Disaster 586
6.2.4 Prism 587
6.2.5 The UW Neutron Therapy Facility 588
6.2.6 Safety by Design: Radiotherapy Machines 589
6.3 System and Data Security 595
6.3.1 Real Life Examples 595
6.3.2 Security Management – the Context 597
6.3.3 Ideas for a Security Program 599
6.4 Summary 601
6.5 Exercises 601

Chapter 7 Epilogue 603
7.1 The Halting Problem—A Metaphoric Version 604
7.2 Where Do New Ideas Come From? 604
7.3 Making A Difference 605

Appendix A Notes on Lisp, Other Languages, and Resources 607
A.1 Lisp Notes 608
A.1.1 CLOS and MOP 609
A.1.2 Optimization 612
A.1.3 Books on Lisp 615
A.1.4 Common Lisp Systems, Libraries, and Resources 616
A.2 Other Programming Languages 617
A.2.1 The `Factorial` Function: A Snapshot Language Comparision 617
A.3 Other Resources 619
A.4 Exercises 620

Bibliography 627

Index 647