

Andreas Holzinger

Biomedical Informatics

Discovering Knowledge in Big Data



 Springer

Andreas Holzinger

Biomedical Informatics

Discovering Knowledge in Big Data

 Springer

Andreas Holzinger
Medical University Graz and
Graz University of Technology
Graz, Austria

Additional material to this book can be downloaded from <http://extra.springer.com>.

ISBN 978-3-319-04527-6 ISBN 978-3-319-04528-3 (eBook)
DOI 10.1007/978-3-319-04528-3
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014932669

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my family and friends

Preface

Amongst all the challenges which medicine and health care in all parts of our world are facing, including exploding costs, finite resources, aging population, etc., maybe one is *the* grand challenge: How to deal with the increasingly large, high-dimensional, weakly structured, and complex data sets and the growing masses of unstructured information produced by modern biomedical sciences. The future trend towards personalized medicine has resulted in an explosion in the amount of generated biomedical data sets from various sources, for example genomics, proteomics, metabolomics, lipidomics, transcriptomics, epigenetics, microbiomics, fluxomics, and phenomics.

This “Big Data” rapidly exceeds the capacity a single medical doctor may be able to handle manually. In the daily routine today, a medical doctor is no longer able to keep pace with these developments and to remember all the thousands of facts of the many diverse patients she or he has. Although human cognitive capacities are excellent in certain areas, in dealing with complex data sets the human learning is limited to a certain extent. This makes the application and further development of sophisticated machine learning approaches a pressing need.

For quite some time now, I have worked persistently on a synergistic combination of methodologies and approaches of two areas, which offer ideal conditions towards solving these challenges towards finding new, efficient, and user-centered algorithms and tools to deal with the “Big Data” challenge: Human–Computer Interaction (HCI) and Knowledge Discovery and Data Mining (KDD), with the goal of supporting human learning with machine learning to interactively discover new, previously unknown *insights into the data*.

Imagine a future where medical doctors would be able to ask questions (e.g., “What are the similarities/differences between patients with symptom X and patients with symptom Y”) to their patient data *and get relevant answers*. The advantageous possibilities are endless if the doctors have an integrated overview on all *relevant* data at their clinical workplace and could find and diagnose diseases well in advance before they might become symptomatically apparent.

This book constitutes the lecture notes to my one-semester course where I try—and one can never be perfect—to provide undergraduate students of biomedical engineering, software engineering, and computer science a broad overview on “Biomedical Informatics” with emphasis on the aforementioned challenges, consequently focusing on a blend of aspects on data, information, and knowledge. Academic freedom lets teach in an egocentric way; consequently, I bring in what I find most interesting and which is amongst my research interests, following a research-based teaching style: I am passionate on extending advanced methods including time (e.g., information entropy) and space (e.g., computational topology), along with user-centered software engineering methods to create interactive software for mobile applications and content analytics techniques, and I follow three promising research streams: Graph-Based Data Mining, Entropy-Based Data Mining, and Topological Data Mining. Naturally, in the limited time of this course I can only scratch the surface and it is impossible to touch each and every interesting aspect—but as a first taster the goal is to open the eyes of the students for the current problems and to encourage them to study further.

Such a foreword is always an opportunity to thank people. This is not an easy task as many people contribute to the development of such a book, either directly or indirectly, and because there are so many, it is always the risk of forgetting somebody, so I will use the plural: I thank my Institutes both at Graz University of Technology and the Medical University of Graz for the academic freedom, the intellectual environment, and the opportunity to teach these fascinating topics to the students. I thank all my colleagues in academia, the clinical domain, and industry for their ongoing and constant fruitful support. I cordially thank all my team and the members from my research group and all my students for their valuable criticism and feedback, and my family and friends for their nurturing encouragement. Last but not least I thank the Springer management and production team for their great and smooth support!

Graz, Austria

Andreas Holzinger

Overview

The Journey Through This Course

The goal of this one-semester course is to provide the students with a broad overview on “Biomedical Informatics” with focus on data, information, and knowledge. We will follow the hypothesis that information can bridge the gap between scientific biomedical research and clinical health practice. The course consists of the following 12 lectures (see Slide 0-1):

Slide 0-1: Roadmap Through This Course

The course consists of the following 12 lectures:

1. **Introduction: Computer Science Meets Life Sciences.** We start with the basics of life sciences, including biochemical and genetic fundamentals, some cell-physiological basics, and a brief overview of the human body; we answer the question “what is biomedical informatics,” and we conclude with an outlook into the future.
2. **Fundamentals of Data, Information, and Knowledge.** In the second lecture we start with a look on data sources, review some data structures, discuss standardization versus structurization, review the differences between data, information, and knowledge, and close with an overview of information entropy.
3. **Structured Data: Coding, Classification (ICD, SNOMED, MeSH, UMLS).** In the third lecture we focus on standardization, ontologies, and classifications, in particular on the International Statistical Classification of Diseases, the Systematized Nomenclature of Medicine, Medical Subject Headings, and the Unified Medical Language.

(continued)

(continued)

4. **Biomedical Databases: Acquisition, Storage, Information Retrieval, and Use.** In the fourth lecture we get a first impression of a hospital information system, we discuss some basics of data warehouse systems and biomedical data banks, and we concentrate on information retrieval.
5. **Semi-structured, Weakly Structured, and Unstructured Data.** In the fifth lecture we review some basics of XML, before we concentrate on network theory and discuss transcriptional regulatory networks, protein–protein networks, and metabolic networks.
6. **Multimedia Data Mining and Knowledge Discovery.** In the sixth lecture we determine types of knowledge, focus on the basics of data mining, and close with text mining and semantic methods, such as Latent Semantic Analysis, Latent Dirichlet Allocation, and Principal Component Analysis.
7. **Knowledge and Decision: Cognitive Science and Human–Computer Interaction.** In the seventh lecture we review the fundamentals of perception, attention, and cognition; discuss the human decision-making process, reasoning, and problem solving; and learn some principles of differential diagnosis and a few basics on human error.
8. **Biomedical Decision Making: Reasoning and Decision Support.** In the eighth lecture we start with the question “Can computers help doctors to make better decisions?” and apply the basics from Lecture 7 to the principles of decision support systems and case-based reasoning systems.
9. **Interactive Information Visualization and Visual Analytics.** In the ninth lecture we start with the basics of visualization science; review some visualization methods, including Parallel Coordinates, Radial Coordinates, and Star Plots; and learn a few things about the design of interactive visualizations.
10. **Biomedical Information Systems and Medical Knowledge Management.** In the tenth lecture we discuss workflow modeling, some basics of business enterprise hospital information systems, Picture Archiving and Communication Systems, and some standards, including DICOM and HL-7.
11. **Biomedical Data: Privacy, Safety, and Security.** In the eleventh lecture we start with the famous IOM “Why do accidents happen?” report and its influence on safety engineering, and concentrate on aspects of data protection and privacy issues of medical data.
12. **Methodology for Information Systems: System Design, Usability, and Evaluation.** Finally in the twelfth lecture we slip into the developer perspective and have a look on design standards, usability engineering methods, and on how we evaluate such systems.

Contents

Lecture 1 Introduction: Computer Science Meets Life Science	1
1 Learning Goals	1
2 Advance Organizer	1
3 Acronyms	3
4 Key Problems	4
5 Our World in Data	4
6 What Is Life?	10
6.1 The Building Blocks of Life	11
6.2 Proteins	13
6.2.1 From Amino Acids to Protein Structures	13
6.2.2 Tertiary Structure of a Protein	14
6.2.3 Protein Analytics	15
6.3 DNA and RNA	18
6.4 Cell Physiology	20
6.5 Organ Systems	21
6.5.1 Tissue	22
6.5.2 Organs	22
6.5.3 Cardiovascular System	23
6.5.4 Anatomical Axes	23
7 What Is Biomedical Informatics?	23
7.1 Medicine Versus Informatics	23
7.2 Computer Machinery	24
7.3 Medical Informatics	30
7.4 Biomedical Informatics	31
8 Future Outlook and Research Avenues	35
8.1 Grand Challenges	35
8.2 Personalized Medicine	36
8.3 Biomarker Discovery	39

9	Exam Questions	40
9.1	Yes/No Decision Questions	40
9.2	Multiple Choice Questions (MCQ)	41
9.3	Free Recall Block	43
10	Answers	46
10.1	Answers to the Yes/No Questions	46
10.2	Answers to the Multiple Choice Questions (MCQ)	47
10.3	Answers to the Free Recall Questions	49
	References	51
	Lecture 2 Fundamentals of Data, Information, and Knowledge	57
1	Learning Goals	57
2	Advance Organizer	57
3	Acronyms	59
4	Key Problems	60
5	Data in the Biomedical Domain	60
5.1	Data Sources in Biomedical Informatics	60
5.2	Levels of Data Structures	61
5.3	Abstract Data Structures	63
5.4	Big Data Pools in the Health Domain	65
5.5	Standardization Versus Structurization	67
5.6	Data Dimensionality	69
5.6.1	Multivariate and Multidimensional	69
5.6.2	Point Cloud Datasets	71
6	A Clinical View on Data, Information, Knowledge	75
7	A Closer Look on Information	78
7.1	What Is Information?	78
7.2	Information Entropy	81
8	Future Outlook	90
9	Exam Questions	94
9.1	Yes/No Questions	94
9.2	Multiple Choice Questions (MCQ)	95
9.3	Free Recall Questions	97
10	Answers	100
10.1	Answers to the Yes/No Questions	100
10.2	Answers to the Multiple Choice Questions (MCQ)	101
10.3	Answers to the Free Recall Questions	103
	References	105
	Lecture 3 Structured Data: Coding and Classification	109
1	Learning Goals	109
2	Advance Organizer	109
3	Acronyms	111
4	Key Problems	111

- 5 Standardization 112
 - 5.1 The Need for Standardization in Medicine 112
 - 5.2 Inaccuracy of Medical Data 113
 - 5.3 Data Standardization 115
- 6 Modeling Biomedical Knowledge 120
- 7 Ontologies 122
 - 7.1 Ontology Languages 127
 - 7.2 OWL 128
- 8 Medical Classifications 129
- 9 Future Outlook 137
- 10 Exam Questions 140
 - 10.1 Yes/No Questions 140
 - 10.2 Multiple Choice Questions (MCQ) 141
 - 10.3 Free Recall Questions 143
- 11 Answers 146
 - 11.1 Answers to the Yes/No Questions 146
 - 11.2 Answers to the Multiple Choice Questions (MCQ) 147
 - 11.3 Answers to the Free Recall Questions 149
- References 151

Lecture 4 Biomedical Databases: Acquisition, Storage, Information Retrieval, and Use 153

- 1 Learning Goals 153
- 2 Advance Organizer 153
- 3 Acronyms 156
- 4 Key Problems 157
- 5 A First View on Hospital Information Systems 158
 - 5.1 Goals and Challenges of Hospital Information Systems 158
 - 5.2 Workflows 160
 - 5.3 Architecture of HIS 161
- 6 Databases 162
 - 6.1 Data Warehouse 164
 - 6.2 Data Marts 164
 - 6.3 Biomedical Databanks and Cloud Computing 165
 - 6.4 Cloud Computing 165
 - 6.5 Biomedical Databases 167
- 7 Information Retrieval 170
 - 7.1 Data Retrieval vs. Information Retrieval 171
 - 7.2 Text Retrieval 172
 - 7.3 IR Process 172
 - 7.4 Formal Notation 172
 - 7.5 Taxonomy of IR Models 175
 - 7.5.1 Set Theoretic Example: Boolean Model 177
 - 7.5.2 Example Algebraic Model: Vector Space Model 178
 - 7.5.3 Example: Probabilistic Model (Bayes' Rule) 182

8	Future Outlook	187
9	Exam Questions	188
	9.1 Yes/No Decision Questions	188
	9.2 Multiple Choice Questions (MCQ)	189
	9.3 Free Recall Block	191
10	Answers	194
	10.1 Answers to the Yes/No Questions	194
	10.2 Answers to the Multiple Choice Questions (MCQ)	195
	10.3 Answers to the Free Recall Questions	197
	References	199
Lecture 5 Semi-structured, Weakly Structured, and Unstructured		
	Data	203
1	Learning Goals	203
2	Advance Organizer	203
3	Acronyms	205
4	Key Problems	205
5	Review on Data	206
	5.1 Well-Structured Data	208
	5.2 Semi-structured Data	209
	5.3 Weakly Structured Data	211
	5.4 On the Topology of Data	212
6	Networks = Graphs + Data	213
	6.1 Networks in Biological Systems	213
	6.2 Network Theory	214
	6.2.1 Basic Concepts of Networks	214
	6.2.2 Computational Graph Representation	214
	6.2.3 Network Metrics	215
	6.2.4 Graphs from Point Cloud Datasets	219
	6.3 Network Examples	225
	6.3.1 The Human Brain as Network	225
	6.3.2 Systems Biology and Human Diseases	225
	6.3.3 Gene Networks	227
	6.4 The Essence: Three Types of Biomedical Networks	228
	6.4.1 Transcriptional Regulatory Networks	228
	6.4.2 Protein–Protein Networks	229
	6.4.3 Metabolic Networks	231
	6.5 Structural Homologies	233
7	Future Outlook	234
8	Exam Questions	235
	8.1 Yes/No Decision Questions	235
	8.2 Multiple Choice Questions (MCQ)	236
	8.3 Free Recall Block	238

- 9 Answers 241
 - 9.1 Answers to the Yes/No Questions 241
 - 9.2 Answers to the Multiple Choice Questions (MCQ) 242
 - 9.3 Answers to the Free Recall Questions 244
- References 247

Lecture 6 Multimedia Data Mining and Knowledge Discovery 251

- 1 Learning Goals 251
- 2 Advance Organizer 251
- 3 Acronyms 253
- 4 Key Problems 253
- 5 Knowledge Discovery 256
 - 5.1 What Is Knowledge? 256
 - 5.2 Implicit vs. Explicit Knowledge 258
 - 5.3 Differences Between KDD and DM 258
- 6 Data Mining Methodologies 260
 - 6.1 Definitions 260
 - 6.2 Example Tasks 261
 - 6.3 Taxonomy of Methods 261
 - 6.4 Supervised Learning 263
 - 6.4.1 Artificial Neural Networks 264
 - 6.4.2 Clinical Example: Model for End-Stage Liver Disease 268
 - 6.4.3 Bayesian Network 271
 - 6.4.4 Alternative Approach: Support Vector Machines 275
 - 6.5 Text Mining and Semantic Methods 276
 - 6.5.1 Latent Semantic Analysis 277
 - 6.5.2 Latent Dirichlet Allocation 278
 - 6.5.3 Principal Components Analysis 279
- 7 Future Outlook 280
- 8 Exam Questions 285
 - 8.1 Yes/No Decision Questions 285
 - 8.2 Multiple Choice Questions (MCQ) 286
 - 8.3 Free Recall Block 288
- 9 Answers 291
 - 9.1 Answers to the Yes/No Questions 291
 - 9.2 Answers to the Multiple Choice Questions (MCQ) 292
 - 9.3 Answers to the Free Recall Questions 293
- References 295

Lecture 7 Knowledge and Decision: Cognitive Science and Human–Computer Interaction 299

- 1 Learning Goals 299
- 2 Advance Organizer 299
- 3 Acronyms 301
- 4 Key Problems 302

- 5 Human Information Processing 302
 - 5.1 Decision Making and Reasoning 302
 - 5.2 The Three-Storage Memory Model 304
 - 5.2.1 Example: Visual and Audial Information Processing 306
 - 5.2.2 Central Executive System 308
 - 5.2.3 Selective Attention 309
 - 5.3 Clinical Decision-Making Process 310
 - 5.4 Reasoning and Problem-Solving Procedures 312
 - 5.4.1 Hypothetico-deductive Model (HDM) vs. PCDA Deming Wheel 312
 - 5.4.2 Signal Detection Theory 316
 - 5.4.3 Differential Diagnosis 319
 - 5.4.4 Rough Set Theory (RST) 320
 - 5.4.5 Heuristic Decision Making 326
- 6 Human Error 327
- 7 Future Outlook 328
- 8 Exam Questions 330
 - 8.1 Yes/No Decision Questions 330
 - 8.2 Multiple Choice Questions (MCQ) 331
 - 8.3 Free Recall Block 333
- 9 Answers 336
 - 9.1 Answers to the Yes/No Questions 336
 - 9.2 Answers to the Multiple Choice Questions (MCQ) 337
 - 9.3 Answers to the Free Recall Questions 339
- References 341

Lecture 8 Biomedical Decision Making: Reasoning and Decision

- Support** 345
 - 1 Learning Goals 345
 - 2 Advance Organizer 345
 - 3 Acronyms 347
 - 4 Key Problems 347
 - 5 Decision Support Systems 348
 - 5.1 Decision Models 349
 - 5.2 Evolution of DSS 350
 - 5.3 Design Principles of DSS 353
 - 5.4 Clinical Guidelines 355
 - 6 Case-Based Reasoning 360
 - 7 Future Outlook 364
 - 8 Exam Questions 365
 - 8.1 Yes/No Decision Questions 365
 - 8.2 Multiple Choice Questions (MCQ) 366
 - 8.3 Free Recall Block 368

9 Answers 371
 9.1 Answers to the Yes/No Questions 371
 9.2 Answers to the Multiple Choice Questions (MCQ) 372
 9.3 Answers to the Free Recall Questions 374
 References 376

Lecture 9 Interactive Information Visualization and Visual

Analytics 379
 1 Learning Goals 379
 2 Advance Organizer 379
 3 Acronyms 381
 4 Key Problems 381
 5 Fundamentals of Visualization Science 382
 5.1 Verbal Information Versus Visual Information 382
 5.2 Is a Picture Really Worth a Thousand Words? 385
 5.3 Informatics as Semiotics Engineering 386
 5.4 Visualization Process 388
 5.5 The Case of John Snow 392
 6 Visualization Methods 393
 6.1 Overview 394
 6.2 Parallel Coordinates 395
 6.3 Radial Coordinate Visualization 399
 6.4 Star Plots 401
 7 Visual Analytics 403
 8 Future Outlook 406
 9 Exam Questions 407
 9.1 Yes/No Decision Questions 407
 9.2 Multiple Choice Questions (MCQ) 408
 9.3 Free Recall Block 410
 10 Answers 413
 10.1 Answers to the Yes/No Questions 413
 10.2 Answers to the Multiple Choice Questions (MCQ) 414
 10.3 Answers to the Free Recall Questions 416
 References 418

Lecture 10 Biomedical Information Systems and Medical

Knowledge Management 421
 1 Learning Goals 421
 2 Advance Organizer 421
 3 Acronyms 423
 4 Key Problems 423
 5 Workflow Modeling 424
 5.1 Workflow and Decision Support 425
 5.2 Formal Modeling 425
 5.3 Example Clinical Workflow 428
 5.4 Workflows in Bioinformatics 432

- 6 Hospital Information Systems 432
 - 6.1 Architectures 433
 - 6.2 Process-Oriented Information Systems 434
- 7 Multimedia in the Hospital 434
 - 7.1 PACS 435
 - 7.2 Data Standards (DICOM, HL7, LOINC) 436
- 8 Future Outlook 442
- 9 Exam Questions 445
 - 9.1 Yes/No Decision Questions 445
 - 9.2 Multiple Choice Questions (MCQ) 446
 - 9.3 Free Recall Block 448
- 10 Answers 451
 - 10.1 Answers to the Yes/No Questions 451
 - 10.2 Answers to the Multiple Choice Questions (MCQ) 452
 - 10.3 Answers to the Free Recall Questions 454
- References 456
- Lecture 11 Biomedical Data: Privacy, Safety, and Security 459**
 - 1 Learning Goals 459
 - 2 Advance Organizer 459
 - 3 Acronyms 461
 - 4 Key Problems 462
 - 5 Standardization and Health Care 463
 - 5.1 What Is Risk? 463
 - 5.2 The IOM Report 464
 - 5.3 Medical Error 466
 - 5.3.1 Eindhoven Classification Model 466
 - 5.3.2 Adverse Event Reporting 467
 - 5.3.3 Human Error 468
 - 5.3.4 Risk Management 468
 - 5.3.5 Ubiquitous Devices 470
 - 5.3.6 Context-Aware Patient Safety 471
 - 5.4 Safety, Security and Technical Dependability 473
 - 6 Patient Data Privacy 475
 - 7 Private Personal Health Record 480
 - 8 Future Outlook 486
 - 9 Exam Questions 487
 - 9.1 Yes/No Decision Questions 487
 - 9.2 Multiple Choice Questions (MCQ) 488
 - 9.3 Free Recall Block 490
 - 10 Answers 493
 - 10.1 Answers to the Yes/No Questions 493
 - 10.2 Answers to the Multiple Choice Questions (MCQ) 494
 - 10.3 Answers to the Free Recall Questions 496
 - References 498

Lecture 12 Methodology for Information Systems: System Design, Usability, and Evaluation 501

1 Learning Goals 501

2 Advance Organizer 501

3 Acronyms 505

4 Key Problems 506

5 A Framework for Understanding Usability 507

6 Standards 509

 6.1 EU Directive: Medical Device Directive 510

 6.2 ISO Standards 511

 6.3 Quality Management Process Cycle 514

 6.4 Software Product Quality Model 515

7 Usability Engineering 516

 7.1 Usability Engineering Methods 516

 7.2 How to Measure Usability? 517

 7.2.1 The System Usability Scale (SUS) 517

 7.2.2 The Software Usability Measurement Inventory (SUMI) 518

 7.2.3 Usability Measurement Metrics 518

 7.3 User-Centered Design and Development 519

8 Evaluation 525

9 Technology Acceptance 526

10 Future Outlook 527

11 Exam Questions 530

 11.1 Yes/No Decision Questions 530

 11.2 Multiple Choice Questions (MCQ) 531

 11.3 Free Recall Block 533

12 Answers 536

 12.1 Answers to the Yes/No Questions 536

 12.2 Answers to the Multiple Choice Questions (MCQ) 537

 12.3 Answers to the Free Recall Questions 539

References 541

Index 547

Index

A

Abduction, 57, 76, 77
Abstraction, 57, 58, 76, 109, 110, 120, 121, 124, 203, 301
Acceptable risk, 459, 469
Accessibility, 58, 63, 66, 73, 129, 137, 162, 169, 253, 254, 283, 393, 464, 475, 483, 501, 502
Accreditation, 422, 437, 501
AceDB, 122, 153
Act, 300, 301, 314, 462, 484, 501, 513, 514, 531, 537
Adverse event, 328, 331, 337, 459, 461, 467, 468, 473, 494
Anonymization, 398, 459, 475, 486, 488, 494
Artifact/surrogate, 58
Artificial neural network (ANN), 177, 251, 264–267, 269, 275, 276, 318
Association rule learning, 251
Audit, 482–484, 501, 502
Authentication, 158, 166, 429, 459, 462, 471, 476, 478, 479

B

Bioinformatics, 1, 11, 19, 20, 63, 136, 165, 168, 210, 211, 214, 233, 379, 399, 421–433
Bioinformatics workflow management system, 421, 432
Biological data visualization, 379
Biological system, 109, 203
Biomarker, 1, 39
Biomedical, 1, 23–25, 27, 29–35, 37, 41, 57, 60, 63, 65, 67, 71, 73, 109, 120–122, 125–127, 135, 136, 151, 153–165, 167, 189, 195, 203, 213, 228, 251, 254, 256,

260, 267–268, 302, 345–376, 386, 396, 401, 405, 421–498

Biomedical data, 1, 31, 33, 34, 60, 459–498
Brute force, 263, 299
Business intelligence (BI), 153, 154, 379, 380
Business process reengineering (BPR), 421

C

Case-based reasoning (CBR), 299, 345, 360–364, 367, 373
Cassandra, 153
Certainty factor model (CF), 345, 352, 365, 366, 371, 372
Certification, 131, 440–442, 502, 512
Cladogram, 153
CLARION, 345
Classical Medicine, 2
Classification, 39, 68, 90, 109–151, 167, 170, 183, 203, 251, 252, 254, 258, 261–263, 266, 276, 282, 284, 295, 300, 318, 323, 357, 358, 360, 379, 405, 418, 423, 462, 466–467
Classification system, 154
Clinical decision support (CDS), 346
Clinical decision support system (CDSS), 35, 276, 346
Clinical pathway, 421
Cloud computing, 154, 157, 165–166, 442
Cluster analysis, 223, 252
Clustering, 3, 4, 10, 14, 64, 154, 156, 165, 170, 171, 187, 216–220, 222, 223, 232, 239, 245, 252, 255, 258, 260–263, 276, 281, 284, 357, 358, 360, 366, 372, 380, 394, 396, 398, 401, 405, 418
Coding, 18, 19, 35, 86, 109–151, 203, 227, 232, 284, 382, 529

Cognition, 11, 31, 261, 299, 305, 380, 381, 393
 Cognitive, 2, 32, 74, 80, 91, 114, 126, 160, 172,
 256, 257, 259, 281, 299–341, 345, 351,
 357, 383, 384, 389, 391, 460, 461, 468,
 502, 517
 Cognitive load, 91, 299
 Cognitive modeling, 327, 334, 340, 502
 Cognitive science, 126, 299–341, 351
 Cognitive walkthrough, 502
 Collective intelligence, 346
 Computerized physician order entry (CPOE),
 3, 30, 41, 47, 154, 156, 189, 408, 414
 Computing, 2, 27–31, 51, 154, 156, 157,
 165–166, 188, 189, 224, 255, 256,
 280, 300, 354, 408, 412, 414, 423,
 425, 428–429, 433, 442–444, 446,
 451, 452, 459, 470, 471, 473, 486,
 493, 498, 504, 541
 Confidentiality, 166, 460, 462, 464, 470, 471,
 475, 486, 512
 Confounding variable, 300
 Consistency, 58, 125, 502, 508, 518
 Consistency inspection, 502
 Content analytics, 281, 379, 380, 388, 392,
 406, 442, 444
 Correlation coefficient, 300
 Crowd sourcing, 346

D

Data, 1, 57, 109, 153, 203, 251, 300, 346, 379,
 421, 459, 504
 Data Mart (DM), 154, 163, 164, 170, 179, 253,
 258–259, 295, 300, 301, 311–312, 317
 Data mining, 2, 74, 92, 170, 172, 206, 223,
 259–295, 317, 333, 339, 351, 354, 355,
 367, 379, 380, 398, 399, 412
 Data model, 110, 171, 204, 210, 403, 407, 413,
 477
 Data protection, 157, 159, 398, 442, 446, 452,
 459, 460, 486
 Data quality, 58, 68, 157, 208
 Data security, 195, 429, 460
 Dataset, 58, 68, 70, 74, 81, 87–89, 115, 170,
 258, 263, 264, 273, 276, 321, 404, 405,
 438, 486
 Data structure, 58, 59, 63, 94, 95, 100, 101,
 123, 223
 Data visualization, 69, 380, 388–390, 393, 398,
 399, 408, 414
 Decision making, 24, 31, 131, 259, 260, 270,
 283, 299, 300, 302–304, 310–312, 315,
 317, 318, 320, 326, 327, 331, 333, 334,

336, 339, 340, 345–376, 381, 407,
 413, 424
 Decision support system (DSS), 35, 276,
 346–360, 366, 367, 370, 372, 375
 Deduction, 58, 76, 77, 301, 349, 351, 353
 Diagnosis, 2, 34, 35, 37, 110, 118, 130, 158,
 204, 232, 299, 300, 302, 303, 312, 317,
 320, 321, 327, 346–349, 353, 359, 365,
 382, 434, 460, 510, 518, 531, 537
 Diagnostic and Statistical Manual (DSM), 110,
 136, 204
 Differential diagnosis (DDX), 299, 301, 303,
 319–320, 330, 336
 Digital imaging and communications in
 medicine (DICOM), 119, 141, 147,
 166, 421, 422, 434
 DIK Model, 58
 DIKW Model, 58
 Dirty data, 58
 Disparity, 58
 Distance matrix method, 154
 DSM. *See* Diagnostic and Statistical Manual
 (DSM)
 DXplain, 346

E

Effectiveness, 167, 254, 502, 508, 513,
 530, 536
 Efficiency, 68, 115, 216, 218, 357, 360, 423,
 424, 502, 504, 508, 509, 515
 Emotion, 300, 502, 527–529
 End user, 28, 31, 34, 41, 47, 58, 73, 116, 135,
 159, 165, 172, 173, 189, 195, 259, 279,
 304, 354, 367, 372, 387, 390, 403, 406,
 408–410, 414, 416, 431, 480, 487, 493,
 501–505, 507, 508, 514, 516, 517,
 519–526, 530, 531, 533, 534, 536,
 537, 539, 540
 End-user programming (EUP), 31, 41, 47, 189,
 502, 531, 537
 EnsEMBL, 15, 22, 41, 47, 95, 101, 154, 165,
 171, 188, 195, 276
 ENTREZ, 154–156, 167
 Entropy, 11, 15, 57–59, 78–90, 95–97, 101,
 102, 222, 294
 Errors, 2, 29, 58, 68, 86, 92, 115, 123, 208, 264,
 265, 275, 299, 302, 303, 317, 326–328,
 330, 331, 336, 337, 359, 429, 430, 459,
 460, 464–474, 487–489, 494, 503,
 504, 508
 Evaluation, 57, 67, 77, 91, 93, 160, 261, 264,
 266, 312, 355, 382, 431, 501–541

Evidence-based medicine (EBM), 3, 24, 37, 112, 115, 140, 146, 161, 178, 300, 347, 422
 Experiment, 58, 81, 82, 90, 168, 169, 525
 Expert system, 346, 347, 350, 351, 353, 357, 365, 367, 371, 373, 443, 463
 Extensible Markup Language (XML), 68, 75, 110, 119, 120, 128, 141, 147, 156, 169, 175, 204, 207, 209–211, 235, 240, 241, 246, 252, 429, 437, 439
 External validity, 300
 Extract, transform, and load (ETL), 154

F

Federated database system, 154, 166
 Formative Evaluation, 503

G

GALEN, 112, 125, 204
 GAMUTS in radiology, 346
 Gedanken, 58, 81, 82
 Genetic algorithm, 154
 Genomics, 1, 2, 12, 20, 32, 60–62, 66, 154, 167, 168, 195, 206, 227, 236, 242, 356
 GOLD, 154, 253, 255
 Graphical user interface (GUI), 505, 524

H

Hadoop, 154, 155
 Hazard, 459, 460, 469
 Hbase, 154
 Health Level Seven (HL 7), 422, 437, 503
 Heart rate variability (HRV), 58, 59
 Heuristic evaluation, 503, 517, 532, 538
 Hospital information system (HIS), 65, 118, 158–162, 168, 195, 208, 209, 347, 366, 372, 406, 421, 422, 428, 429, 432–434, 436, 437, 440, 442, 444–446, 448, 451, 452, 454, 464, 476
 HRV. *See* Heart rate variability (HRV)
 Human fallibility, 328, 460, 468, 487, 493
 Hypothetico-deductive model (HDM), 299, 300, 312–315

I

ICD. *See* International Classification of Diseases (ICD)
 ILLIAD, 346
 Induction, 58, 76, 77, 94, 100, 263, 320, 325, 345, 353

Information entropy, 58, 81–90
 Information Extraction (IE), 155
 Information overload, 58
 Information quality, 31–33, 58, 433, 501
 Information retrieval (IR), 93, 110, 127, 153–199, 204, 251, 527–528
 Information visualization, 91, 93, 379–418
 Integrating the Healthcare Enterprise (IHE), 422, 442, 443
 Internal validity, 300
 International Classification of Diseases (ICD), 68, 109, 110, 115, 130–132, 136, 140, 141, 146, 147, 203, 204, 232, 423
 IR. *See* Information retrieval (IR)
 ISO 9001, 501, 503, 512
 ISO 9241, 503, 531, 537
 ISO 9241-10, 503
 ISO 9241-11, 503
 ISO 13407, 503, 519, 523, 525, 530, 531, 536, 537
 ISO 13485 (2003), 503, 512, 514
 ISO 14971 (2007), 461, 503, 512
 ISO 62304 (2006), 503
 ISO/HL7, 503
 ISO/IEEE, 503
 ISO/OECD, 503

K

k-Anonymity, 398, 460
 Knowledge, 2, 57, 109, 153, 203, 209, 299, 345, 380, 431, 460, 502
 Knowledge Discovery (KD), 7, 35, 36, 73, 74, 92, 159, 170, 172, 187, 251–295, 347, 354, 388, 391, 398, 425, 444
 Knowledge Extraction, 252, 357, 358

L

Large data Multidimensional, 59, 91, 95, 101, 165, 170, 251, 252, 264, 405, 442
 Learnability, 277, 503, 508, 515
 Learning curve, 503

M

MapReduce, 154, 155
 Mashup, 31, 155, 480–482, 487, 493
 Mash-up, 41, 47, 187, 502–504, 531, 537
 Medical Classification, 119, 129–137, 204
 Medical error, 299, 327, 328, 460, 464–474, 488, 494
 Medical safety design, 504

- Medicine, 2, 3, 19, 23–24, 30, 32, 36–39, 109–112, 130, 132–133, 140, 146, 203–205, 234, 304, 327, 346, 347, 398, 422, 436, 465, 473, 474
- Medizin Produkte Gesetz (MPG), 501, 504, 513
- MEDLINE, 136, 155, 465
- Memorability, 504, 508
- Mental model, 79, 80, 95, 101, 166, 380, 385, 391, 407, 408, 413, 414, 502, 504
- MeSH, 109, 110, 130, 133–136, 140, 146, 203, 204
- Metabolomics, 2, 66
- Metadata, 110, 155, 204
- Metadata Model, 110, 204
- Methodology, 85, 91, 110, 166, 173, 187, 204, 260–280, 284, 327, 330, 334, 336, 340, 485, 501–541
- MMMDB, 155
- Molecular, 2, 9–13, 15, 16, 19, 23, 31, 34, 36–38, 40, 41, 46, 47, 60, 155, 168, 169, 188, 189, 194, 195, 206, 225, 227, 229, 230, 233, 234, 236, 238, 242, 244, 261, 350, 379, 385, 408, 414, 437
- Multidimensional, 59, 63, 67, 69, 74, 93, 94, 100, 110, 204, 279, 380, 382, 395, 406
- Multidimensional scaling, 380
- Multimedia, 62, 115, 251–295, 383, 421, 434–439, 441
- Multi-modality, 59
- Multi-variate, 33, 59, 380, 383
- MYCIN, 345, 346, 350–353, 365, 366, 371, 372
- N**
- National Electrical Manufacturers Association (NEMA) Paradigm, 111, 119, 422, 438
- Natural language processing (NLP), 115, 155, 172, 176, 253, 281
- Neural networks, 155, 177, 195, 251, 262, 264–265, 267–268, 270, 284, 366, 372
- Nomen Nescio (N.N) patient safety, 460
- Non-relational database, 155
- Nosography, 110, 204
- Nosology, 110, 204
- O**
- Omics data, 2, 36, 38, 65–67, 76, 163, 164
- Online Mendelian Inheritance in Man (OMIM), 136, 155, 167
- Ontology, 68, 110, 117, 122–130, 136, 138, 140, 141, 146, 147, 185, 204, 207, 232, 388
- Ontology engineering, 110, 204
- P**
- Parallel coordinates, 93, 215, 380, 394–398, 411, 415, 417
- Participatory design, 504
- PCA. *See* Principal component analysis (PCA)
- PDB. *See* Protein database (PDB)
- Perception, 301
- Performance, 2, 3, 26–27, 31, 32, 112, 113, 153, 158, 171, 179, 186, 259, 264, 284, 299, 304, 309, 312, 314, 318, 356, 442, 443, 473, 474, 502, 504, 508, 513, 518, 529, 536
- Performance measure, 504
- Personally identifying information prevention, 460
- Pervasive, 2, 27–29, 31, 37–39, 41, 47, 189, 195, 408, 414, 541
- Pervasive Health, 2, 28, 37
- P-Health Model, 2, 38–39
- Phylogenetics, 155, 167
- Picture Archiving and Communication System (PACS), 422
- Plan-Do-Check-Act (PDCA), 300–301, 312, 314, 514
- Principal Component Analysis (PCA), 252, 263, 277, 279, 280
- Privacy, 460
- Privacy policy, 460, 482–484
- PROSITE, 155
- Protected health information (PHI), 461, 462
- Protein database (PDB), 5, 6, 35, 69, 155, 167, 168, 195
- Proteome, 2, 12, 16
- Proteomics, 1, 2, 15, 16, 32, 60–62, 66, 95, 101, 167, 261
- Pseudonymization, 459, 461, 475–480, 488, 494
- Q**
- Qualitative research, 301
- Quantitative research, 301
- Quasi-Identifiers, 398, 460, 461
- R**
- RadViz, 379, 380, 394, 399–401, 407, 408, 411, 413, 414, 417
- Reasoning, 59, 76, 77, 104, 120, 125–127, 129, 138, 161, 172, 173, 175, 176, 195, 258,

- 299–304, 311–327, 330, 334, 336, 340,
345–376, 380, 388, 403, 408, 414,
471, 472
- Receiver-operating characteristic (ROC), 276,
301, 318, 331, 337
- Relational database, 122, 155
- S**
- Safety, 157, 159, 423, 442, 444, 446, 452,
459–498, 504, 506, 509–511, 513,
518, 527
- Safety-critical systems research, 461
- Safety engineering, 459, 461
- Safety risk management, 461
- Satisfaction, 430, 504, 507, 508, 518, 527, 530,
536
- Security, 37, 157, 159, 195, 423, 429, 442–444,
446, 452, 459–498, 509–512, 515, 537
- Semiotic engineering, 380, 386, 387
- Semiotics, 380, 386–388, 408, 414, 504
- Semi-structured data, 122, 207, 209–211,
237, 242
- Sensitive data, 398, 461
- Similarity table, 156
- SNOMED, 109–111, 130, 132–133, 136, 140,
141, 146, 151, 185, 203–205
- SNOP, 20, 110, 132
- Software usability measurement inventory
(SUMI), 505, 506, 518
- Software usability scale (SUS), 505, 506, 517,
518, 531, 537
- Space, 2, 4, 14, 40, 46, 63, 70–72, 74, 86, 94,
100, 135, 160, 172, 176–182, 188, 207,
212, 223, 231, 233, 238, 242, 244, 266,
275–278, 280, 380, 381, 393–396, 399,
401, 404, 406, 408, 409, 414, 508
- Spatiality, 59
- SQL, 156, 163
- SRS, 156
- Star plot, 379, 380, 395, 401, 403, 408, 410,
414, 416
- Structural complexity, 59, 222
- Supervised learning, 252, 262–276
- Supervised learning algorithm, 252
- Support vector machine (SVM), 251, 252, 262,
270, 275–276, 318
- Swiss-Cheese model, 328, 461
- SWISS-PROT, 156, 169
- Symbolic reasoning, 301
- System features, 110, 204
- T**
- Task analysis, 505, 520
- Technological Performance, 26, 31
- Terminology, 68, 110, 115, 123, 124, 127, 129,
132, 136, 184, 204, 437
- Thinking aloud, 505, 517, 519, 522, 531, 537
- Time dependency Voxel, 59
- Translational Medicine, 3
- Triage, 301, 311, 330, 336
- U**
- Unified Language System (UMLS), 109, 110,
127, 130, 133, 135–137, 140, 141, 146,
147, 151, 203
- UniGene, 156
- Unsupervised learning, 252, 262, 263
- Usability engineering, 93, 488, 494, 501, 504,
505, 516–524, 530, 534, 540
- User interface (UI), 35, 36, 92, 172, 351, 353,
366, 372, 381, 403, 405, 472, 502, 503,
505, 516, 524
- V**
- Validation, 269, 354, 403, 440, 501, 505, 520,
525, 530, 536
- Verification, 157, 261, 501, 505, 520, 525
- Visual analytics, 379–381, 388, 392, 403–406,
410, 416
- Visualization, 8, 9, 74, 91–93, 135, 171, 174,
215, 259, 261, 268, 284, 379–417, 432,
442, 443, 504
- Visualization mantra, 381, 404, 408, 414
- Von Neumann Computer, 3
- W**
- Workflow, 118, 158–161, 164, 347, 353, 362,
421, 422, 424–434, 436, 438, 441, 507,
516, 521