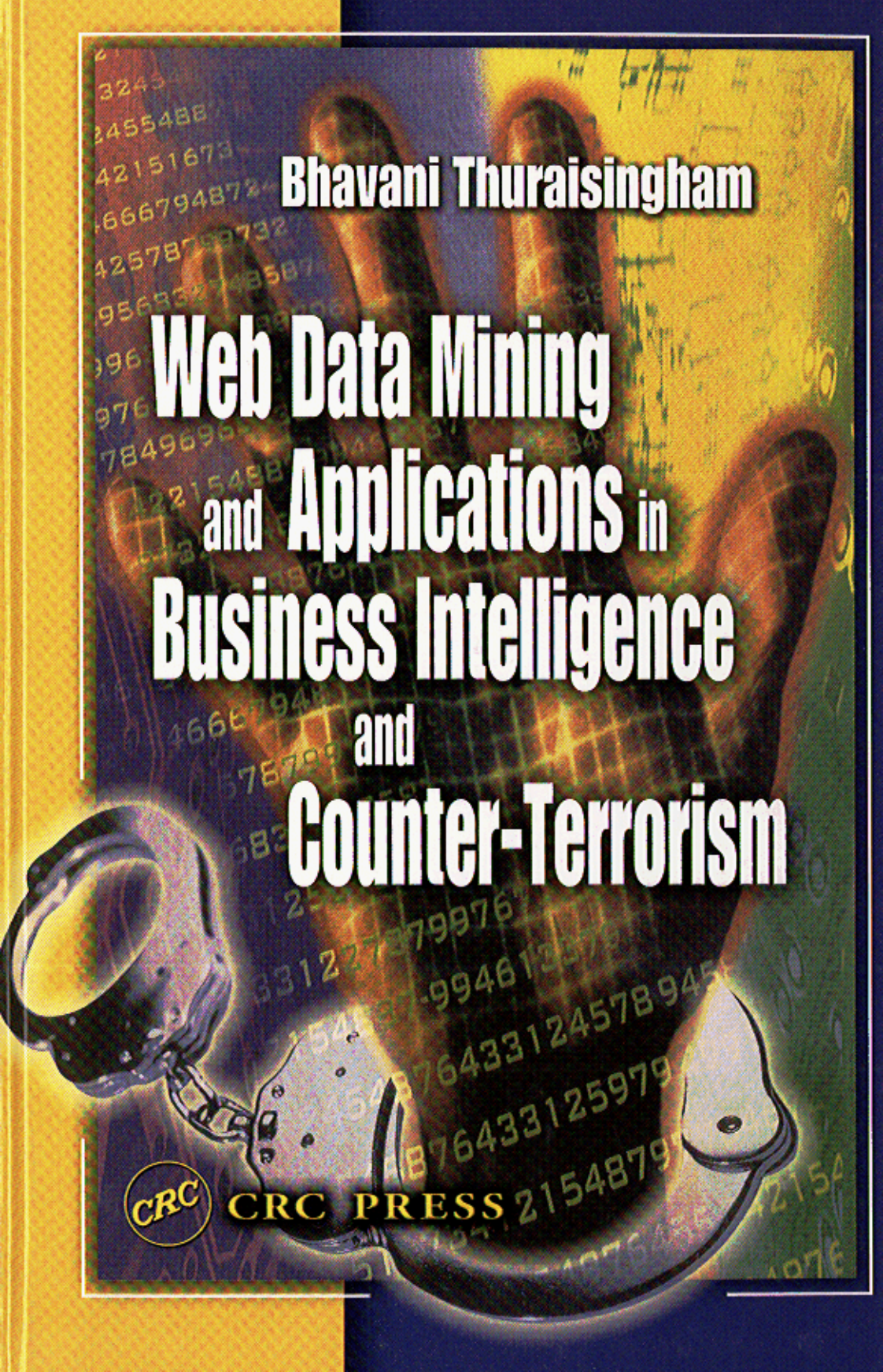


The background of the book cover features a close-up of a human hand with fingers slightly spread, set against a dark, textured background. Overlaid on the hand and background are numerous strings of binary code (0s and 1s) in a light, glowing green or yellow color. The overall aesthetic is high-tech and digital.

Bhavani Thuraisingham

**Web Data Mining
and Applications in
Business Intelligence
and
Counter-Terrorism**



CRC PRESS

Contents

Preface	xv
1 Introduction	1
Introduction	1
Supporting Technologies for Web Data Mining	3
Web Data Mining Concepts	6
Web Data Mining, E-Commerce, and Business Intelligence	7
Web Data Mining and Counter-Terrorism	8
Organization of this Book	9
How Do We Proceed?	13
PART I SUPPORTING TECHNOLOGIES FOR WEB DATA MINING ...	15
Introduction to Part I	16
2 The World Wide Web, E-Commerce, and Business Intelligence	17
Introduction	17
Evolution of the Web	18
Introduction to E-Commerce	21
Relationship to Web Data Mining	31
Summary	33
3 Data Mining	35
Introduction	35
Data Mining Technologies	36
Concepts and Techniques in Data Mining	38
Directions and Trends in Data Mining	39
Relationship to Web Mining	41
Summary	41
4 Core Data Mining Technologies	43
Introduction	43
Data Management and Data Warehousing	44
Statistical Reasoning	51
Machine Learning	53

	Visualization	54
	Parallel Processing	56
	Decision Support	57
	Architectural Support for Data Mining	58
	Relationship to Web Data Mining	65
	Summary	65
5	Web Database Management	67
	Introduction	67
	Web Databases	68
	Semistructured Databases	74
	Metadata, Ontologies, and the Web	75
	Distributed, Heterogeneous, and Legacy Databases on the Web	78
	Data Warehousing on the Web	82
	Architectural Aspects	83
	Relationship to Web Data Mining	87
	Summary	88
6	Information Retrieval Systems	89
	Introduction	89
	Text Retrieval	89
	Image Retrieval	91
	Video Retrieval	92
	Audio Retrieval	94
	Multimedia Data and Information Management	94
	Question Answering Systems	98
	Markup Languages	99
	Relationship to Web Data Mining	100
	Summary	101
7	Information Management Technologies	103
	Introduction	103
	Collaboration and Data Management	103
	Knowledge Management	106
	Agents for the Web	109
	Training and Distance Learning	115
	Wireless Information Management and Pervasive Computing	115
	Sensor Information Management	116
	Quality of Service Aspects	117
	Some Directions	118
	Relationship to Web Data Mining	119
	Summary	120
8	The Semantic Web	121
	Introduction	121
	Semantic Web Concepts	122

RDF.....	124
Role of Ontologies	126
Agents and the DAML Program	127
Semantic Web as a Database.....	128
XML, RDF, and Interoperability.....	129
Web Services.....	130
Note on E-Commerce, Business Intelligence, and the Semantic Web	131
Web versus the Semantic Web	132
Data Mining and the Semantic Web	133
Summary	133
Conclusion to Part I	135

PART II WEB DATA MINING: TECHNIQUES, TOOLS, AND

TRENDS

Introduction to Part II	138
-------------------------------	-----

9 Data Mining and the Web

Introduction	139
Mining Data on the Web.....	140
Mining Usage Patterns.....	144
Web Structure Mining	145
Applications and Directions	145
Summary	148

10 Processes and Techniques for Web Data Mining.....

Introduction	149
Process of Web Data Mining	149
Web Data Mining Outcomes, Approaches, and Techniques.....	159
Web Data Mining versus Data Mining	166
Summary	167

11 Mining Databases on the Web.....

Introduction	169
Concepts in Web Database Mining.....	169
Mining Semistructured Databases	173
Metadata and Web Mining.....	178
Mining Distributed, Heterogeneous, Legacy, and Federated Databases on the Web	182
Architectures and Web Data Mining	190
Summary	192

12 Information Retrieval and Web Data Mining.....

Introduction	195
Search Engines and Web Data Mining	195

Multimedia Data Mining and the Web	200
Some Other Aspects.....	216
Summary	218
13 Information Management and Web Data Mining.....	219
Introduction.....	219
Collaborative Data Mining.....	219
Knowledge Management and Web Data Mining	221
Training and Web Data Mining	222
Agents and Web Data Mining	223
Wireless Computing and Web Data Mining	224
Sensor Web Mining.....	226
Quality of Service and Web Data Mining	226
Other Aspects	228
Summary	229
14 Semantic Web Mining	231
Introduction.....	231
Concepts in Semantic Web Mining.....	231
XML, RDF, and Web Data Mining.....	232
Ontologies and Web Data Mining	234
Agents and Web Data Mining	235
Web Mining and the Semantic Web as a Database.....	236
Semantic Interoperability and Web Mining	238
Web Services and Web Mining	239
Web Mining vs. Semantic Web Mining	240
A Note on E-Commerce, Business Intelligence, and Semantic Web Mining	240
Summary	242
15 Mining Usage Patterns and Structure on the Web.....	243
Introduction.....	243
Web Usage Mining Outcomes and Techniques.....	244
Web Usage Mining Analysis.....	245
CRM and Business Intelligence Applications	247
Mining Structure on the Web	252
Summary	252
16 Prototypes, Products, and Standards for Web Data Mining	255
Introduction.....	255
Prototypes and Products for Data Mining.....	256
Web Data Mining Products	261
Standards for Web Mining	268
Summary	273

17 Some Applications for Web Mining	275
Introduction	275
E-Commerce and E-Business	275
Business Intelligence	277
Customer Relationship Management	278
Marketing and Sales	279
Enterprise Resource Management	280
Manufacturing and Planning	281
Education and Training	282
Telecommunications	283
Financial	283
Physical Sciences, Social Sciences, and Engineering	284
Medicine and Biotechnology	285
Counter-Terrorism	286
Summary	286
Conclusion to Part II	288
PART III WEB DATA MINING FOR COUNTER-TERRORISM	289
Introduction to Part III	290
18 Some Information on Terrorism, Security Threats, and Protection Measures	291
Introduction	291
Natural Disasters and Human Errors	292
Noninformation-Related Terrorism	294
Information-Related Terrorism	296
Biological, Chemical, and Nuclear Attacks	301
Attacks on Critical Infrastructures	302
Nonreal-Time Threats vs. Real-Time Threats	303
Aspects of Counter-Terrorism	304
A Note on Privacy	310
Summary	311
19 Web Data Mining for Counter-Terrorism	313
Introduction	313
Web Data Mining for Counter-Terrorism	314
Analyzing the Techniques	319
Link Analysis	322
Summary	324
20 Mining Web Databases for Counter-Terrorism	327
Introduction	327
Web Database Mining	328
Semistructured Database Mining	333
Metadata Mining	333

Warehouse Mining	334
Distributed and Heterogeneous Database Mining	336
Other Aspects	341
Summary	341
21 Information Retrieval and Web Mining for Counter-Terrorism.....	343
Introduction.....	343
Exploiting and Using Intelligent Search Engines.....	344
Applying Multimedia Data Mining	346
Other Aspects	354
Summary	356
22 Information Management and Web Mining for Counter-Terrorism.....	357
Introduction.....	357
Collaborative Data Mining.....	357
Knowledge Management and Web Mining.....	358
Training and Web Mining.....	359
Agents and Web Mining.....	360
Wireless Information Management and Web Mining.....	361
Sensor Information Management and Web Mining.....	362
Quality of Service and Web Mining	363
Some Other Aspects.....	365
Summary	365
23 Semantic Web Mining for Counter-Terrorism	367
Introduction.....	367
Semantic Web Mining for Counter-Terrorism	367
Building the Semantic Web and Web Mining for Counter-Terrorism	371
Summary	374
24 Web Usage and Structure Mining for Counter-Terrorism.....	375
Introduction.....	375
Web Usage Mining for Counter-Terrorism.....	375
Web Structure Mining for Counter-Terrorism.....	377
Summary	379
25 National Security, Privacy, Civil Liberties, and Web Mining	381
Introduction.....	381
Background of the Inference Problem	383
Mining, Warehousing, and Inference	384
Inductive Logic Programming and Inference	386
Privacy Issues	388

Inference Problem and Privacy	390
Privacy-Enhanced Data Mining	391
Civil Liberties vs. National Security	391
Summary	393
26 Revisiting Security Threats with Respect to Web Mining	395
Introduction	395
Natural Disasters, Human Error, and Malicious Attacks	395
Noninformation-Related Terrorism	396
Information-Related Terrorism	397
Biological, Chemical, and Nuclear Attacks	397
Attacks on Critical Infrastructures	398
Nonreal-Time Threats vs. Real-Time Threats	399
Revisiting Privacy	399
Summary	400
27 E-Commerce, Business Intelligence, and Counter-Terrorism	401
Introduction	401
Models for Counter-Terrorism	402
Architectures for Counter-Terrorism	403
Functions for Counter-Terrorism	406
Revisiting Business Intelligence	408
Summary	409
Conclusion to Part III	412
28 Summary and Directions	413
Introduction	413
Summary of the Book	413
Challenges and Directions for Web Data Mining	419
Challenges and Directions for Web Data Mining for Counter-Terrorism	420
Impact of the Department of Homeland Security	422
Where Do We Go from Here?	423
PART IV APPENDIXES	427
A Data Management Systems: Developments and Trends	429
Developments in Database Systems	430
Status, Vision, and Issues	434
Data Management Systems Framework	435
Building Information Systems from the Framework	438
Relationship between the Texts	441
Summary	441
References	443

B	Database Systems and Related Technologies.	445
	Relational and Entity-Relationship Data Models	445
	Architectural Issues	447
	Database Design	449
	Database Administration	450
	Database Management System Functions	451
	Distributed Databases	458
	Heterogeneous Database Integration.	460
	Federated Databases	461
	Client/Server Databases	462
	Migrating Legacy Databases and Applications	465
	Data Warehousing.	465
	Object Technology	467
	Summary	472
	References	473
C	Data and Information Security	477
	Access Control and Other Security Concepts.	478
	Secure Systems	479
	Secure Database Systems	480
	Emerging Trends.	481
	Impact of the Web.	483
	Summary	484
	References	484
D	References	487
	About the Author	495
	Index	499