

Application of **Big Data for National Security**



A Practitioner's Guide to Emerging Technologies

BIBLIOTHEQUE DU CERIST

```
Content-Type: application/soap+xml; charset=utf-8
Content-Length: 3932
<?xml version="1.0"?>
<soap:Envelope soap:encodingStyle="">
  <soap:Body xmlns:m="http://192.168.1.1/loc">
    <m:SecurityArray>
      <m>PasswordIn>*****</m>PasswordIn>
    </m:SecurityArray>
  </soap:Body>
```



Babak Akhgar, Gregory B. Saathoff, Hamid R. Arabnia,
Richard Hill, Andrew Staniforth, and Petra Saskia Bayerl



Application of Big Data for National Security

A Practitioner's Guide to Emerging Technologies

Edited by

Babak Akhgar

Gregory B. Saathoff

Hamid R. Arabnia

Richard Hill

Andrew Staniforth

Petra Saskia Bayerl



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Butterworth-Heinemann is an imprint of Elsevier



Contents

List of Contributors	xv
About the Editors	xvii
Foreword by Lord Carlile of Berriew	xix
Preface by Edwin Meese III	xxi
Acknowledgments	xxiii

SECTION 1 INTRODUCTION TO BIG DATA

CHAPTER 1 An Introduction to Big Data	3
What Is Big Data?	3
How Different Is Big Data?	4
More on Big Data: Types and Sources	4
Structured Data	5
Unstructured Data	5
Semi-Structured Data	5
The Five V's of Big Data	5
Volume	6
Velocity	6
Variety	6
Veracity	7
Value	7
Big Data in the Big World	7
Importance	7
Advantages and Applications	7
Analytical Capabilities of Big Data	9
Data Visualization	9
Greater Risk Intelligence	9
Satisfying Business Needs	9
Predictive Modeling and Optimization	10
Streaming Analytics	10
Identifying Business Use Cases	10
Video and Voice Analytics	10
Geospatial Analytics	10
An Overview of Big Data Solutions	11
Google BigQuery	11
IBM InfoSphere BigInsights	11
Big Data on Amazon Web Services	11
Clouds for Big Data	11
Conclusions	12
References	12

CHAPTER 2	Drilling into the Big Data Gold Mine: Data Fusion and High-Performance Analytics for Intelligence Professionals	14
	Introduction	14
	The Age of Big Data and High-Performance Analytics	14
	Technology Challenges	15
	Building the Complete Intelligence Picture	16
	Examples	19
	Scenario 1: Fusion and Michigan State Police	19
	Scenario 2: National Security and Intelligence Solution in the Middle East	19
	Conclusion	20
	Reference	20

SECTION 2 CORE CONCEPTS AND APPLICATION SCENARIOS

CHAPTER 3	Harnessing the Power of Big Data to Counter International Terrorism.....	23
	Introduction	23
	A New Terror	24
	Fertilizer Plot	24
	International Dimension	25
	Executive Action	26
	Vulnerabilities Emerge	27
	Assessing the Threat	27
	Suicide Terror	29
	Joining the Dots	30
	Held to Account	30
	Strategic Approach	32
	Changing Threat Landscape	33
	Embracing Big Data	34
	Conclusion	36
	References	37
CHAPTER 4	Big Data and Law Enforcement: Advances, Implications, and Lessons from an Active Shooter Case Study	39
	The Intersection of Big Data and Law Enforcement	39
	Case Example and Workshop Overview	41
	Situational Awareness	43
	Looking into the Past	43
	Interacting with the Public	44
	Alerting and Prediction	44
	Twitter as a Social Media Source of Big Data	45
	Social Media Data Analyzed for the Workshop	45

Tools and Capabilities Prototypes During the Workshop	46
Word Cloud Visualization	46
Dynamic Classification of Tweet Content	46
Content-Based Image Retrieval	47
Maximizing Geographic Information	48
Detecting Anomalies	49
Influence and Reach of Messaging	49
Technology Integration	50
Law Enforcement Feedback for the Sessions	51
Discussion	51
Acknowledgments	52
References	52
CHAPTER 5 Interpretation and Insider Threat: Rereading the Anthrax Mailings of 2001 Through a “Big Data” Lens.....	55
Introduction	55
Importance of the Case	57
The Advancement of Big Data Analytics After 2001	58
Relevant Evidence	59
Potential for Stylometric and Sentiment Analysis	61
Potential for Further Pattern Analysis and Visualization	63
Final Words: Interpretation and Insider Threat	64
References	65
CHAPTER 6 Critical Infrastructure Protection by Harnessing Big Data	68
Introduction	68
What Is a CI System?	68
Understanding the Strategic Landscape into Which Big Data Must Be Applied	69
What Is Meant by an Overarching Architecture?	73
The SCR	73
Underpinning the SCR	76
Strategic Community Architecture Framework	77
Conclusions	80
References	80
CHAPTER 7 Military and Big Data Revolution.....	81
Risk of Collapse	81
Into the Big Data Arena	82
Simple to Complex Use Cases	83
Canonic Use Cases	87
Filtering	88
Correlation of Data Over Space and Time	88

More on the Digital Version of the Real World (See the World as Events)	89
Quality of Data, Metadata, and Content	90
Real-Time Big Data Systems	91
Application Principles and Constraints	91
Logical View	93
Implementing the Real-Time Big Data System	95
Batch Processing (into the “Batch Layer”)	95
Processing Layer (into the “Batch Layer”)	95
Spark	96
Data Stream Processing (into the “Streaming Layer”)	97
Alerts and Notifications (into the “Publishing Layer”)	98
Filtering Processing Fitting in Memory (into the “Streaming Layer”)	98
Machine Learning and Filtering (into the “Batch and Streaming Layers”)	99
Online Clustering (into the “Streaming Layer”)	100
Results Publication	100
Build the Layers	101
Insight into Deep Data Analytics Tools and Real-Time Big Data Systems	102
Add Fault Tolerance	103
Security	103
Adding Flexibility and Adaptation	104
Very Short Loop and Battlefield Big Data Datacenters	104
Conclusions	104
Further Reading	106
CHAPTER 8 Cybercrime: Attack Motivations and Implications for Big Data and National Security	108
Introduction	108
Defining Cybercrime and Cyberterrorism	110
Attack Classification and Parameters	111
Who Perpetrates These Attacks?	113
Script Kiddies	113
Web Defacers	114
Hackers	114
Pirates	114
Phone Phreakers	115
Tools Used to Facilitate Attacks	115
Motivations	117
Attack Motivations Taxonomy	118
Political	118
Ideological	120
Commercial	120

Emotional	120
Informational/Promotional	121
Financial	121
Personal	121
Exploitation	122
Detecting Motivations in Open-Source Information	122
Conclusion	123
References	123

SECTION 3 METHODS AND TECHNOLOGICAL SOLUTIONS

CHAPTER 9 Requirements and Challenges for Big Data Architectures 131

What Are the Challenges Involved in Big Data Processing?	131
Deployment Concept	131
Technological Underpinning	132
The Core Technologies	132
Planning for a Big Data Platform	134
Infrastructure Requirements	134
Capacity Planning Considerations	137
Cloud Computing Considerations	137
Conclusions	139
References	139

CHAPTER 10 Tools and Technologies for the Implementation of Big Data 140

Introduction	140
Techniques	141
Representation, Storage, and Data Management	141
Analysis	142
A/B Testing	142
Association Rule Learning	143
Classification	143
Crowdsourcing	143
Data Mining	143
Natural Language Processing and Text Analysis	143
Sentiment Analysis	144
Signal Processing	144
Visualization	144
Computational Tools	144
Hadoop	145
MapReduce	145
Apache Cassandra	145

Implementation 145
 Implementation Issues 146
 Project Initiation and Launch 146
 Information Technology Project Reference Class 148
 Mitigating Factors 149
 User Factors and Change Management 149
 Data Sources and Analytics 150
 Cloud/Crowd sourcing 150
 Corporate Systems 150
 Analytics Philosophy: Analysis or Synthesis 151
 Governance and Compliance 152
 Data Protection Requirements and Privacy 152
 References 153

CHAPTER 11 Mining Social Media: Architecture, Tools, and Approaches to Detecting Criminal Activity 155

Introduction 155
 Mining of Social Networks for Crime 157
 Text Mining 158
 Natural Language Methods 158
 Symbolic Approach 158
 Statistical Approach 159
 Connectionist Approach 159
 General Architecture and Various Components of Text Mining 159
 Lexical Analysis 159
 POS Tagging 160
 Parsing 160
 Named Entity Recognition 161
 Co-reference Resolution 161
 Relation Extraction 161
 Concept Extraction 162
 Topic Recognition 163
 Sentiment Analysis 163
 Semantic Analysis 163
 Machine Translation 163
 Bayesian Networks 163
 Automatic Extraction of BNs from Text 165
 Dependence Relation Extraction from Text 165
 Variables Identification 166
 BN Structure Definition 166
 Probability Information Extraction 166

Aggregation of Structural and Probabilistic Data	166
BNs and Crime Detection	167
General Architecture	167
Example of BN Application to Crime Detection: Covert Networks	169
Conclusions	169
References	170

CHAPTER 12 Making Sense of Unstructured Natural Language Information..... 173

Introduction	173
Big Data and Unstructured Data	174
Aspects of Uncertainty in Sense Making	175
Situation Awareness and Intelligence	176
Situation Awareness: Short Timelines, Small Footprint	176
Intelligence: Long(er) Timelines, Larger Footprint	176
Processing Natural Language Data	177
Structuring Natural Language Data	178
Two Significant Weaknesses	179
Ignoring Lexical Clues on Credibility and Reliability	179
Out of Context, Out of Mind	180
An Alternative Representation for Flexibility	180
Conclusions	182
References	182

CHAPTER 13 Literature Mining and Ontology Mapping Applied to Big Data 184

Introduction	184
Background	185
Parameter Optimized Latent Semantic Analysis	186
Improving the Semantic Meaning of the POLSA Framework	186
Web Services	186
ARIANA: Adaptive Robust Integrative Analysis for Finding Novel Associations	187
Conceptual Framework of ARIANA	187
Ontology Mapping	189
Data Stratification and POLSA	189
Relevance Model	190
Reverse Ontology Mapping	192
Visualization and Interface	192
Implementation of ARIANA for Biomedical Applications	193
OM and MGD Creation	194
Creation of the MGD	195
Data Stratification and POLSA	195
Parameter Optimized Latent Semantic Analysis	198

Relevance Model 198
 Reverse Ontology Mapping and I&V 200
 Case Studies 201
 Case Study I: KD: Lethal Drug Interaction 201
 Case Study II: Data Repurposing: AD Study 202
 Discussion 202
 Conclusions 204
 Acknowledgment 205
 References 205

CHAPTER 14 Big Data Concerns in Autonomous AI Systems 209

Introduction 209
 Artificially Intelligent System Memory Management 210
 Sensory Memories 210
 Short-term Artificial Memories 211
 Long-term Artificial Memories 211
 Artificial Memory Processing and Encoding 212
 Short-term Artificial Memory Processing 212
 Long-term Artificial Memory Processing 216
 Implicit Biographical Memory Recall/Reconstruction Using Spectral
 Decomposition Mapping 217
 Constructivist Learning 218
 Adaptation of Constructivist Learning Concepts for Big Data in an AIS 220
 Practical Solutions for Secure Knowledge Development in Big
 Data Environments 221
 Practical Big Data Security Solutions 221
 Optimization of Sociopolitical-Economic Systems and Sentiment Analysis 223
 Conclusions 224
 References 225

SECTION 4 LEGAL AND SOCIAL CHALLENGES

CHAPTER 15 The Legal Challenges of Big Data Application in Law Enforcement 229

Introduction 229
 Attractions of Big Data 229
 Dilemmas of Big Data 230
 Legal Framework 230
 Human Rights 231
 Purpose Limitation and Further Processing 233
 Public Trust and Confidence 234

Conclusions	236
How Far Should Big Data Principles Such as “Do Not Track” and “Do Not Collect” Be Applicable to LEAs, Either in Qualified Format or at All?	236
References	237
CHAPTER 16 Big Data and the Italian Legal Framework: Opportunities for Police Forces	238
Introduction	238
European Legal Framework	239
Directive 95/46/EC and Revision Process Started in 2012	239
Data Retention Directive	241
The Italian Legal Framework	242
Authority for Personal Data Protection	242
The Italian Privacy Code	242
Focus on Italian Police Forces	243
Police Data Processing and Privacy	244
Opportunities and Constraints for Police Forces and Intelligence	245
References	248
CHAPTER 17 Accounting for Cultural Influences in Big Data Analytics	250
Introduction	250
Considerations from Cross-Cultural Psychology for Big Data Analytics	251
Cultural Dependence in the Supply and Demand Sides of Big Data Analytics	252
Cultural Dependence on the Supply Side (Data Creation)	253
Cultural Dependence on the Demand Side (Data Interpretation)	254
(Mis)Matches among Producer, Production, Interpreter, and Interpretation Contexts	256
Integrating Cultural Intelligence into Big Data Analytics: Some Recommendations	257
Conclusions	258
References	259
CHAPTER 18 Making Sense of the Noise: An ABC Approach to Big Data and Security	261
How Humans Naturally Deal with Big Data	261
The Three Stages of Data Processing Explained	262
Stage 1: Reflexive	263
Stage 2: Pre-attentive	263
Stage 3: Attentive	264
The Public Order Policing Model and the Common Operational Picture	265
Applications to Big Data and Security	267
Level 1: Reflexive Response	268

Level 2: Pre-attentive Response268
Level 3: Attentive Response and the Focused, Intellectual Management
of Data269
Application to Big Data and National Security270
A Final Caveat from the FBI Bulletin272
References272

Glossary275
Index279