Contents

About This Book xi
About The Authors xv
Acknowledgments xvii
Chapter 1 Introduction to Text Analytics 1
Overview of Text Analytics 1
Text Mining Using SAS Text Miner 5
Information Retrieval 7
Document Classification 8
Ontology Management 9
Information Extraction 10
Clustering 11
Trend Analysis 12
Enhancing Predictive Models Using Exploratory Text Mining 13
Sentiment Analysis 14
Emerging Directions 15
Handling Big (Text) Data 15
Voice Mining 16
Real-Time Text Analytics 16
Summary 16
References 17
Chapter 2 Information Extraction Using SAS Crawler 19
Introduction to Information Extraction and Organization 19
SAS Crawler 20
SAS Search and Indexing 20
SAS Information Retrieval Studio Interface 20
Web Crawler 22
Breadth First 23
Depth First 24
Web Crawling: Real-World Applications and Examples 24
Understanding Core Component Servers 26
Proxy Server 26
Pipeline Server 27
Component Servers of SAS Search and Indexing 28
Indexing Server 28
Query Server 28

Query Web Server 29
Query Statistics Server 29
SAS Markup Matcher Server 29
Summary 39
References 39
Chapter 3 Importing Textual Data into SAS Text Miner 41
An Introduction to SAS Enterprise Miner and SAS Text Miner 41
Data Types, Roles, and Levels in SAS Text Miner 42
Creating a Data Source in SAS Enterprise Miner 43
Importing Textual Data into SAS 48
Importing Data into SAS Text Miner Using the Text Import Node 49
%TMFILTER Macro 57
Importing XLS and XML Files into SAS Text Miner 58
Managing Text Using SAS Character Functions 62
Summary 67
References 68
Chapter 4 Parsing and Extracting Features 69
Introduction 69
Tokens and Words 70
Lemmatization 70
POS Tags 71
Parsing Tree 71
Text Parsing Node in SAS Text Miner 73
Stemming and Synonyms 73
Identifying Parts of Speech 78
Using Start and Stop Lists 81
Spell Checking 84
Entities 86
Building Custom Entities Using SAS Contextual Extraction Studio 88
Summary 90
References 90
Chapter 5 Data Transformation 93
Introduction 93
Zipf's Law 94
Term-By-Document Matrix 96
Text Filter Node 97
Frequency Weightings 98
Term Weightings 98
Filtering Documents 102
Concept Links 106
Summary 108
References 108

Chapter 6 Clustering and Topic Extraction 111
Introduction 111
What Is Clustering? 111
Singular Value Decomposition and Latent Semantic Indexing 113
Topic Extraction 122
Scoring 130
Summary 130
References 131
Chapter 7 Content Management 133
Introduction 133
Content Categorization 134
Types of Taxonomy 136
Statistical Categorizer 139
Rule-Based Categorizer 141
Comparison of Statistical versus Rule-Based Categorizers 144
Determining Category Membership 145
Concept Extraction 146
Contextual Extraction 150
CLASSIFIER Definition 150
SEQUENCE and PREDICATE_RULE Definitions 155
Automatic Generation of Categorization Rules Using SAS Text Miner 157
Differences between Text Clustering and Content Categorization 159
Summary 160
Appendix 161
References 162
Chapter 8 Sentiment Analysis 163
Introduction 163
Basics of Sentiment Analysis 164
Challenges in Conducting Sentiment Analysis 165
Unsupervised versus Supervised Sentiment Classification 165
SAS Sentiment Analysis Studio Overview 166
Statistical Models in SAS Sentiment Analysis Studio 167
Rule-Based Models in SAS Sentiment Analysis Studio 172
SAS Text Miner and SAS Sentiment Analysis Studio 175
Summary 176
References 177
Case Studies 179
Case Study 1 Text Mining SUGI/SAS Global Forum Paper Abstracts to Reveal Trends 181
Introduction 181
Data 181
Results 189
Trends 190

Summary 194
Instructions for Accessing the Case Study Project 194
Case Study 2 Automatic Detection of Section Membership for SAS Conference Paper Abstract Submissions 197
Introduction 197
Objective 198
Step-by-Step Instructions 198
Summary 208
Case Study 3 Features-based Sentiment Analysis of Customer Reviews 209
Introduction 209
Data 209
Text Mining for Negative App Reviews 210
Text Mining for Positive App Reviews 217
NLP Based Sentiment Analysis 219
Summary 225
Case Study 4 Exploring Injury Data for Root Causal and Association Analysis 227
Introduction 227
Objective 227
Data Description 227
Step-by-Step Instructions 228
Part 1: SAS Text Miner 228
Part 2: SAS Enterprise Content Categorization 234
Summary 238
Case Study 5 Enhancing Predictive Models Using Textual Data 241
Data Description 241
Step-by-Step Instructions 241
Summary 249
Case Study 6 Opinion Mining of Professional Drivers' Feedback 251
Introduction 251
Data 251
Analysis Using SAS® Text Miner 251
Analysis Using the Text Rule-builder Node 258
Summary 272
Case Study 7 Information Organization and Access of Enron Emails to Help Investigation 273
Introduction 273
Objective 273
Step-by-Step Software Instruction with Settings/Properties 274
Summary 281
Case Study 8 Unleashing the Power of Unified Text Analytics to Categorize Call Center Data 283
Introduction 283
Data Description 284

Examining Topics 285
Merging or Splitting Topics 288
Categorizing Content 288
Concept Map Visualization 289
Using PROC DS2 for Deployment DEPLOYMENT 292
Integrating with SAS® Visual Analytics 293
Summary 294
Case Study 9 Evaluating Health Provider Service Performance Using Textual Responses 297
Introduction 297
Summary 311
Index 313