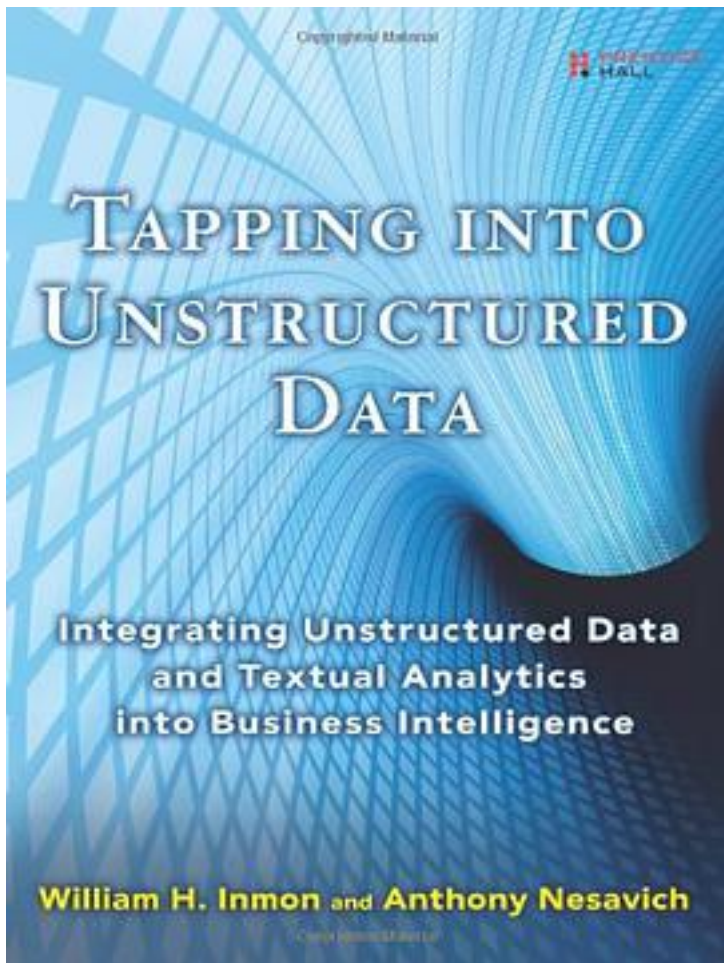


Tapping into Unstructured Data



[Tapping into Unstructured Data_下载链接1](#)

著者:Inmon, William H./ Nesavich, Anthony

出版者:

出版时间:2007-11

装帧:

isbn:9780132360296

"The authors, the best minds on the topic, are breaking new ground. They show how every organization can realize the benefits of a system that can search and present complex ideas or data from what has been a mostly untapped source of raw data."

--Randy Chalfant, CTO, Sun Microsystems The Definitive Guide to Unstructured Data Management and Analysis--From the World's Leading Information Management Expert A wealth of invaluable information exists in unstructured textual form, but organizations have found it difficult or impossible to access and utilize it. This is changing rapidly: new approaches finally make it possible to glean useful knowledge from virtually any collection of unstructured data. William H. Inmon--the father of data warehousing--and Anthony Nesavich introduce the next data revolution: unstructured data management. Inmon and Nesavich cover all you need to know to make unstructured data work for your organization. You'll learn how to bring it into your existing structured data environment, leverage existing analytical infrastructure, and implement textual analytic processing technologies to solve new problems and uncover new opportunities. Inmon and Nesavich introduce breakthrough techniques covered in no other book--including the powerful role of textual integration, new ways to integrate textual data into data warehouses, and new SQL techniques for reading and analyzing text. They also present five chapter-length, real-world case studies--demonstrating unstructured data at work in medical research, insurance, chemical manufacturing, contracting, and beyond. This book will be indispensable to every business and technical professional trying to make sense of a large body of unstructured text: managers, database designers, data modelers, DBAs, researchers, and end users alike. Coverage includes

- *What unstructured data is, and how it differs from structured data
- *First generation technology for handling unstructured data, from search engines to ECM--and its limitations
- *Integrating text so it can be analyzed with a common, colloquial vocabulary: integration engines, ontologies, glossaries, and taxonomies
- *Processing semistructured data: uncovering patterns, words, identifiers, and conflicts
- *Novel processing opportunities that arise when text is freed from context
- *Architecture and unstructured data: Data Warehousing 2.0
- *Building unstructured relational databases and linking them to structured data
- *Visualizations and Self-Organizing Maps (SOMs), including Compudigm and Raptor solutions
- *Capturing knowledge from spreadsheet data and email
- *Implementing and managing metadata: data models, data quality, and more

William H. Inmon is founder, president, and CTO of Inmon Data Systems. He is the father of the data warehouse concept, the corporate information factory, and the government information factory. Inmon has written 47 books on data warehouse, database, and information technology management; as well as more than 750 articles for trade journals such as Data Management Review, Byte, Datamation, and ComputerWorld. His b-eye-network.com newsletter currently reaches 55,000 people. Anthony Nesavich worked at Inmon Data Systems, where he developed multiple reports that successfully query unstructured data.

Preface xvii 1
 Unstructured Textual Data in the Organization 1 2 The Environments of Structured Data and Unstructured Data 15 3 First Generation Textual Analytics 33 4 Integrating Unstructured Text into the Structured Environment 47 5 Semistructured Data 73 6 Architecture and Textual Analytics 83 7 The Unstructured Database 95 8 Analyzing a Combination of Unstructured Data and Structured Data 113 9 Analyzing Text Through Visualization 127 10 Spreadsheets and Email 135 11 Metadata in Unstructured Data 147 12 A Methodology for Textual Analytics 163 13 Merging Unstructured Databases into the Data Warehouse 175 14 Using SQL to Analyze Text 185 15 Case Study--Textual Analytics in Medical Research 195 16 Case Study--A Database for Harmful Chemicals 203 17 Case Study--Managing Contracts Through an Unstructured Database 209 18 Case Study--Creating a Corporate Taxonomy (Glossary) 215 19 Case Study--Insurance Claims 219 Glossary 227 Index 233

作者介绍:

目录:

[Tapping into Unstructured Data 下载链接1](#)

标签

评论

[Tapping into Unstructured Data 下载链接1](#)

书评

[Tapping into Unstructured Data 下载链接1](#)