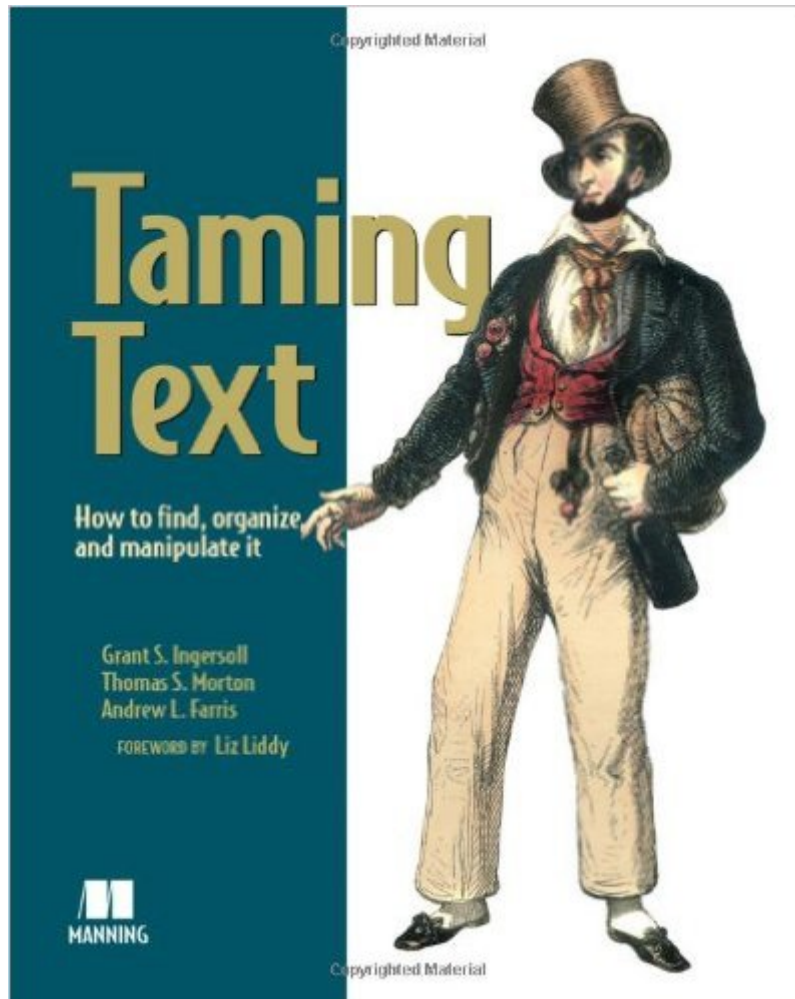# The book was found

Taming Text: How To Find, Organize, And Manipulate It

Synopsis

SummaryTaming Text, winner of the 2013 Jolt Awards for Productivity, is a hands-on, example-driven guide to working with unstructured text in the context of real-world applications. This book explores how to automatically organize text using approaches such as full-text search, proper name recognition, clustering, tagging, information extraction, and summarization. The book guides you through examples illustrating each of these topics, as well as the foundations upon which they are built.About this BookThere is so much text in our lives, we are practically drowningin it. Fortunately, there are innovative tools and techniquesfor managing unstructured information that can throw thesmart developer a much-needed lifeline. You'll find them in thisbook.Taming Text is a practical, example-driven guide to working withtext in real applications. This book introduces you to useful techniques like full-text search, proper name recognition,clustering, tagging, information extraction, and summarization.You'll explore real use cases as you systematically absorb thefoundations upon which they are built.Written in a clear and concise style, this book avoids jargon, explainingthe subject in terms you can understand without a backgroundin statistics or natural language processing. Examples arein Java, but the concepts can be applied in any language.Written for Java developers, the book requires no prior knowledge of GWT. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book. Winner of 2013 Jolt Awards: The Best Booksâ "one of five notable books every serious programmer should read.What's InsideWhen to use text-taming techniquesImportant open-source libraries like Solr and MahoutHow to build text-processing applicationsAbout the AuthorsGrant Ingersoll is an engineer, speaker, and trainer, a Lucenecommitter, and a cofounder of the Mahout machine-learning project. Thomas Morton is the primary developer of OpenNLP and Maximum Entropy. Drew Farris is a technology consultant, software developer, and contributor to Mahout,Lucene, and Solr."Takes the mystery out of verycomplex processes."â "From the Foreword by Liz Liddy, Dean, iSchool, Syracuse UniversityTable of ContentsGetting started taming textFoundations of taming textSearchingFuzzy string matchingIdentifying people, places, and thingsClustering textClassification, categorization, and taggingBuilding an example question answering systemUntamed text: exploring the next frontier

Book Information

Customer Reviews

Text processing is a complex topic and this book makes it more palatable. The unique thing about
this book, that I found valuable, is that it brings together a slew of innovative technologies like
SOLR, Mahout, Tika, etc. The book starts small, building up various related facets, and eventually
culminates into an application that is both complex and applicable in real world situations. I feel
much more at ease dealing with the complexities of text processing after reading this book and
subsequent related research. I would like to thank the authors for providing such a great resource.I
used the companion code to follow along using a Linux Mint VM in Windows 8.

Solr and Lucene has made text search a commodity today. Web applications can plug in fairly
feature-rich search functionality using Solr in a matter of days, and most applications today consider
search functionality a routine must-have. As long as your needs are met with Solr's extensive
feature set, you need look no further. However, search is constantly evolving, and the focus today is
on making search results even more useful to the user. This book addresses this trend, exploring
various strategies and software tools that can be used in conjunction with Solr and Lucene to mine
the text in your indexes to do this. Tools covered are OpenNLP, Carrot and Apache Mahout. The
book builds up towards building a simple question-answering system that relies on search.
OpenNLP is used for named entity extraction (identifying people, places, things, etc) from the
corpus and for categorization. Carrot's Lingo algorithm is used for online clustering of search
results, and Apache Mahout is used to demonstrate clustering and classification solutions.You will
find the book most useful if you work with Solr, Lucene and Java and are looking to get beyond
basic search, since all the tools described fall in the Java search ecosystem. The authors cover

Lucene analysis and tokenization, along with the basic NLP and machine learning theory that are needed, so knowledge of these is not a pre-requisite (although it can help).

I love this book. I also love Python. My first into into Natural Language Processing (NLP) was with NLTK for Python. After getting my hands dirty and starting to get some real work done I realized that I had to move to Java. This book is a great intro into that world. The chapters are organized well into tasks that you probably will need to do at some point when working with text. Chapter 4 is worth the price of this book alone.The examples are easy to follow and the theory is clearly explained.I was already familiar with OpenNLP when I read this book so I cannot comment on how it feels as a beginner.

This book is down-to-earth with perfect balance of depth of the subject and practical applications. I have intermediate level NLP experience, but this book still delivers at this level with easy to read style despite the very technical basis of the topics.

Great book.It would be better, if like Programming Collective Intelligence, it summarizes strengths/weaknesses before the details.

The author does a fantastic job writing about what might otherwise be a dry subject, providng just enough technical information to be useful and allow you to sink your teeth into it.

Good read. A decent starting point if you are just getting into a lot of data/text work

Simple and clear, application oriented. Although I would like it to explain a little more about setting up the tools that are used through the book.

Taming Text: How to Find, Organize, and Manipulate It Time to Get Things Done: Beat Procrastination, Stop Being Lazy, Take Actions, and Master Your Life in 24 Hours (2nd Edition) (Organize Yourself, Organize ... Self Organization, To Do List Book 6) Seek and Find Can You Find Me? (Seek and Find Book) Interactive 3-D Maps: American History: Easy-to-Assemble 3-D Maps That Students Make and Manipulate to Learn Key Facts and Concepts—in a Kinesthetic Way! Social Media: Master, Manipulate, and Dominate Social Media Marketing With Facebook, Twitter, YouTube, Instagram and LinkedIn (Social Media, Social Media ... Twitter, Youtube, Instagram,

Pinterest) Social Media: Master, Manipulate, and Dominate Social Media Marketing Facebook, Twitter, YouTube, Instagram, and LinkedIn This Is Your Brain on Parasites: How Tiny Creatures Manipulate Our Behavior and Shape Society The Hacked World Order: How Nations Fight, Trade, Maneuver, and Manipulate in the Digital Age Brandwashed: Tricks Companies Use to Manipulate Our Minds and Persuade Us to Buy The Perfectly Trained Parrot: Fun and Positive Methods for Taming, Socializing, Trick Training, Release and Solving Behavior Problems 30 Days to Taming Your Tongue: What You Say (and Don't Say) Will Improve Your Relationships The Taming of the Queen (The Plantagenet and Tudor Novels) The Taming of Romanticism: European Literature and the Age of Biedermeier (Harvard Studies in Comparative Literature) The Taming of the Shrew (Shakespeare for Everyone) Taming of the Shrew: Sixty-Minute Shakespeare Series The Taming of the Drew Kidnapped: The Taming of the Princess Bitch Taming the Paper Tiger at Home The Taming of the Queen Chasing Hope: Taming Natasha, Luring a Lady (The Stanislaskis)