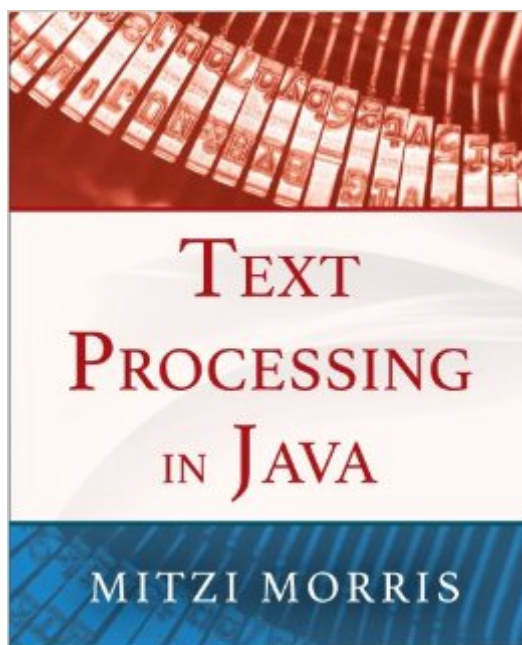


The book was found

# Text Processing In Java



## Synopsis

This book teaches you how to master the subtle art of multilingual text processing and prevent text data corruption. It provides an introduction to natural language processing using Lucene and Solr. It gives you tools and techniques to manage large collections of text data, whether they come from news feeds, databases, or legacy documents. Each chapter contains executable programs that can also be used for text data forensics. Topics covered:

- Unicode code points
- Character encodings from ASCII and Big5 to UTF-8 and UTF-32LE
- Character normalization using International Components for Unicode (ICU)
- Java I/O, including working directly with zip, gzip, and tar files
- Regular expressions in Java
- Transporting text data via HTTP
- Parsing and generating XML, HTML, and JSON
- Using Lucene 4 for natural language search and text classification
- Search, spelling correction, and clustering with Solr 4

Other books on text processing presuppose much of the material covered in this book. They gloss over the details of transforming text from one format to another and assume perfect input data. The messy reality of raw text will have you reaching for this book again and again.

## Book Information

Paperback: 328 pages

Publisher: Colloquial Media Corporation (January 1, 2014)

Language: English

ISBN-10: 0988208725

ISBN-13: 978-0988208728

Product Dimensions: 7.5 x 0.7 x 9.2 inches

Shipping Weight: 1 pounds (View shipping rates and policies)

Average Customer Review: 4.7 out of 5 stars [See all reviews](#) (3 customer reviews)

Best Sellers Rank: #1,212,586 in Books (See Top 100 in Books) #100 in [Books > Computers & Technology > Computer Science > AI & Machine Learning > Natural Language Processing](#) #301 in [Books > Computers & Technology > Internet & Social Media > Online Searching](#) #1235 in [Books > Computers & Technology > Programming > Languages & Tools > Java](#)

## Customer Reviews

This is an essential reference book for Java developers working with text data. It covers in detail the arcane world of character encodings and ends with how to setup Lucene/SOLR search engines. Java's character and string classes are thoroughly explained. There is coverage of Java I/O, regular expressions and common web formats like HTML, JSON etc as well. The code is clear with worked

examples based on the Federalist papers which is pretty cool.

This is a very solid book that explains some of the techniques that I had been lacking, or just confused about. I recommend it as a staple for the Java developer.

Great source for java text processing

[Download to continue reading...](#)

Java: The Ultimate Guide to Learn Java and Python Programming (Programming, Java, Database, Java for dummies, coding books, java programming) (HTML, ... Developers, Coding, CSS, PHP) (Volume 3) JAVA: JAVA in 8 Hours, For Beginners, Learn Java Fast! A Smart Way to Learn Java, Plain & Simple, Learn JAVA Programming Language in Easy Steps, A Beginner's Guide, Start Coding Today! Java: The Simple Guide to Learn Java Programming In No Time (Programming, Database, Java for dummies, coding books, java programming) (HTML, Javascript, Programming, Developers, Coding, CSS, PHP) (Volume 2) Text Processing in Java Java Programming for Kids: Learn Java Step By Step and Build Your Own Interactive Calculator for Fun! (Java for Beginners) Information Processing with Evolutionary Algorithms: From Industrial Applications to Academic Speculations (Advanced Information and Knowledge Processing) Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition Deep Learning: Natural Language Processing in Python with Recursive Neural Networks: Recursive Neural (Tensor) Networks in Theano (Deep Learning and Natural Language Processing Book 3) Deep Learning: Natural Language Processing in Python with GLoVe: From Word2Vec to GLoVe in Python and Theano (Deep Learning and Natural Language Processing) Deep Learning: Natural Language Processing in Python with Word2Vec: Word2Vec and Word Embeddings in Python and Theano (Deep Learning and Natural Language Processing Book 1) Natural Language Processing with Java and LingPipe Cookbook Text Processing with Ruby: Extract Value from the Data That Surrounds You Programming Perl: Unmatched power for text processing and scripting Gregg College Keyboarding & Document Processing (GDP); Lessons 1-120, main text Gregg College Keyboarding & Document Processing (GDP); Lessons 61-120 text Gregg College Keyboarding & Document Processing (GDP); Lessons 1-20 text Python Text Processing with NLTK 2.0 Cookbook CORBA and Java: Where Distributed Objects Meet the Web (Java Masters) Learning Java by Building Android Games - Explore Java Through Mobile Game Development Java Artificial Intelligence: Made Easy, w/ Java Programming; Learn to Create your \* Problem Solving \* Algorithms! TODAY! w/ Machine Learning & Data ... engineering, r programming,

iOS development)

[Dmca](#)