O'REILLY®

2nd Edition

Mining the Social Web

DATA MINING FACEBOOK, TWITTER, LINKEDIN, GOOGLE+, GITHUB, AND MORE

Matthew A. Russell

O'REILLY®

Mining the Social Web

How can you tap into the wealth of social web data to discover who's making connections with whom, what they're talking about, and where they're located? With this expanded and thoroughly revised edition, you'll learn how to acquire, analyze, and summarize data from all corners of the social web, including Facebook, Twitter, LinkedIn, Google+, GitHub, email, websites, and blogs.

- Employ IPython Notebook, the Natural Language Toolkit, NetworkX, and other scientific computing tools to mine popular social websites
- Apply advanced text-mining techniques, such as clustering and TF-IDF, to extract meaning from human language data
- Bootstrap interest graphs from GitHub by discovering affinities among people, programming languages, and coding projects
- Build interactive visualizations with D3.js, an extraordinarily flexible HTML5 and JavaScript toolkit
- Take advantage of more than two-dozen Twitter recipes, presented in O'Reilly's popular "problem/solution/discussion" cookbook format

The example code for this unique data science book is maintained in a public GitHub repository. It's designed to be easily accessible through a turnkey virtual machine that facilitates interactive learning with an easy-to-use collection of IPython Notebooks.

Matthew Russell, Chief Technology Officer at Digital Reasoning Systems, Principal at Zaffra, is a computer scientist who is passionate about data mining, open source, and creating technology to amplify human intelligence.

> "Mining insights through an API is an essential skill to have, whether or not you consider yourself a programmer. This book exposes you to a breadth of key information sources while using tools that make the coding easily accessible."
>
> **—Kevin Makice**
> author of *Twitter API: Up and Running*

> "This book offers all readers a fresh perspective of social web data through illustrative and concise code—all within in the comfort of a web browser! Readers get a fantastic tour of computer science concepts by example, including algorithmic complexity, natural language processing, and the future of the Internet of Things."
>
> **—Jason Yee**
> Data Scientist at Digital Reasoning

DATA/DATA MINING

US $44.99 CAN $47.99
ISBN: 978-1-449-36761-9

54499
9 781449 367619

O'REILLY®

Strata

Making Data Work

Learn how to turn data into decisions.

From startups to the Fortune 500, smart companies are betting on data-driven insight, seizing the opportunities that are emerging from the convergence of four powerful trends:

- New methods of collecting, managing, and analyzing data
- Cloud computing that offers inexpensive storage and flexible, on-demand computing power for massive data sets
- Visualization techniques that turn complex data into images that tell a compelling story
- Tools that make the power of data available to anyone

Get control over big data and turn it into insight with O'Reilly's Strata offerings. Find the inspiration and information to create new products or revive existing ones, understand customer behavior, and get the data edge.

O'REILLY®

Visit oreilly.com/data to learn more.

SECOND EDITION

Mining the Social Web

Matthew A. Russell

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

Mining the Social Web, Second Edition

by Matthew A. Russell

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://my.safaribooksonline.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Editor: Mary Treseler
Production Editor: Kristen Brown
Copyeditor: Rachel Monaghan
Proofreader: Rachel Head
Indexer: Lucie Haskins
Cover Designer: Karen Montgomery
Interior Designer: David Futato
Illustrator: Rebecca Demarest

October 2013: Second Edition

Revision History for the Second Edition:

2013-09-25: First release

See *http://oreilly.com/catalog/errata.csp?isbn=9781449367619* for release details.

While every precaution has been taken in the preparation of this book, the publisher and author assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

ISBN: 978-1-449-36761-9

[LSI]

If the ax is dull and its edge unsharpened, more strength is needed,
but skill will bring success.

—Ecclesiastes 10:10

Table of Contents