INSTANT

Short | Fast | Focused

Web Scraping with Java

Build simple scrapers or vast armies of Java-based bots to untangle and capture the Web

Ryan Mitchell

[PACKT] PUBLISHING

Instant Web Scraping with Java

Build simple scrapers or vast armies of Java-based bots to untangle and capture the Web

Ryan Mitchell

[PACKT]
PUBLISHING

BIRMINGHAM - MUMBAI

Instant Web Scraping with Java

First published: August 2013

Production Reference: 1230813

Published by Packt Publishing Ltd.
Livery Place
35 Livery Street
Birmingham B3 2PB, UK.

ISBN 978-1-84969-688-3

www.packtpub.com

Credits

Author

Ryan Mitchell

Reviewers

Benjamin Hill

Sumant Kumar Raja

Acquisition Editor

Mary Nadar

Commissioning Editors

Sharvari Tawde

Ameya Sawant

Technical Editor

Akashdeep Kundu

Project Coordinator

Joel Goveya

Proofreader

Jonathan Todd

Graphics

Abhinash Sahu

Production Coordinator

Conidon Miranda

Cover Work

Conidon Miranda

Cover Image

Sheetal Aute

About the Author

Ryan Mitchell has ten years of programming experience, including Java, C, Perl, PHP, and Python. In addition to "traditional" programming, she specializes in web technologies, with three years of Drupal development experience, and is Sitecore developer certified.

She graduated from Olin College of Engineering and is currently studying at the Harvard University Extension School for a Masters in Software Engineering.

In addition to academic life, she currently works at Velir Studios as a Web Systems Analyst, and has also worked as a developer for Harvard University and Abine Inc.

About the Reviewers

Benjamin Hill is a product manager working in online and mobile applications. His background includes everything from a bootstrapped crowdsourcing startup in 2006 to companies as big as Comcast and eBay. During that time he has learned a variety of techniques to automate websites that don't yet provide APIs.

Sumant Kumar Raja is an entrepreneur, integration architect, and independent consultant with seven years of experience working for global clients. He has worked on a wide range of complex technical environments including J2EE and SAP Process Integration. He is a Certified ScrumMaster.

He is director of Denarit Consulting Ltd. UK. In past, he has worked in various roles with Accenture for clients in the UK and the U.S.

I am thankful to my family for their unconditional support.

I would like to extend my thanks to Packt Publishing for giving me the opportunity to review this book.

---

www.PacktPub.com

Support files, eBooks, discount offers and more

You might want to visit `www.PacktPub.com` for support files and downloads related to your book.

Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at `www.PacktPub.com` and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at `service@packtpub.com` for more details.

At `www.PacktPub.com`, you can also read a collection of free technical articles, sign up for a range of free newsletters and receive exclusive discounts and offers on Packt books and eBooks.

`http://PacktLib.PacktPub.com`

Do you need instant solutions to your IT questions? PacktLib is Packt's online digital book library. Here, you can access, read and search across Packt's entire library of books.

Why Subscribe?

- Fully searchable across every book published by Packt
- Copy and paste, print and bookmark content
- On demand and accessible via web browser

Free Access for Packt account holders

If you have an account with Packt at `www.PacktPub.com`, you can use this to access PacktLib today and view nine entirely free books. Simply use your login credentials for immediate access.

Table of Contents