

php|architect's Guide to Web Scraping



[php|architect's Guide to Web Scraping 下载链接1](#)

著者:Matthew Turland

出版者:Marco Tabini & Associates, Inc.

出版时间:2010-12

装帧:Paperback

isbn:9780981034515

Despite all the advancements in web APIs and interoperability, it's inevitable that, at some point in your career, you will have to "scrape" content from a website that was not built with web services in mind. And, despite its sometimes less-than-stellar reputation, web scraping is usually an entire legitimate activity—for example, to capture data from an old version of a website for insertion into a modern CMS. This book, written by scraping expert Matthew Turland, covers web scraping techniques and topics that range from the simple to exotic using a variety of technologies and frameworks: • Understanding HTTP requests • The PHP HTTP streams wrapper • cURL • pecl_http • PEAR:HTTP • Zend_Http_Client • Building your own scraping

library • Using Tidy • Analyzing code with the DOM, SimpleXML and XMLReader extensions • CSS selector libraries • PCRE pattern matching • Tips and Tricks • Multiprocessing / parallel processing

作者介绍:

目录:

[php|architect' s Guide to Web Scraping_ 下载链接1_](#)

标签

PHP

Programming

评论

[php|architect' s Guide to Web Scraping_ 下载链接1_](#)

书评

[php|architect' s Guide to Web Scraping_ 下载链接1_](#)