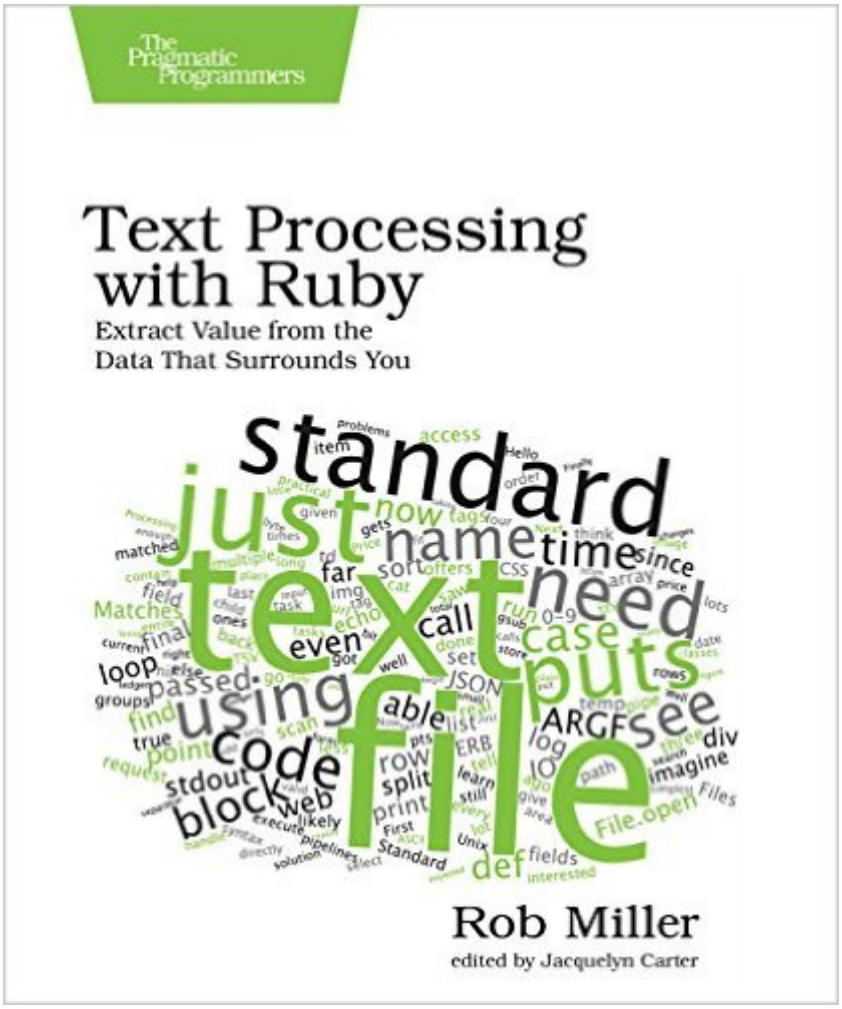
Text Processing With Ruby: Extract Value From The Data That Surrounds You



Synopsis

Text is everywhere. Web pages, databases, the contents of files--for almost any programming task you perform, you need to process text. Cut even the most complex text-based tasks down to size and learn how to master regular expressions, scrape information from Web pages, develop reusable utilities to process text in pipelines, and more.Most information in the world is in text format, and programmers often find themselves needing to make sense of the data hiding within. It might be to convert it from one format to another, or to find out information about the text as a whole, or to extract information fromit. But how do you do this efficiently, avoiding labor-intensive, manual work?Text Processing with Ruby takes a practical approach. You'll learn how to get text into your Ruby programs from the file system and from user input. You'll process delimited files such as CSVs, and write utilities that interact with other programs in text-processing pipelines. Decipher character encoding mysteries, and avoid the pain of jumbled characters and malformed output.You'll learn to use regular expressions to match, extract, and replace patterns in text. You'll write a parser and learn how to process Web pages to pull out information from even the messiest of HTML.Before long you'll be able to tackle even the most enormous and entangled text with ease, scything through gigabytes of data and effortlessly extracting the bits that matter.What You Need:This book requires a passing familiarity with the Ruby programming language, and assumes that you already have Ruby installed on your computer.

Book Information

Paperback: 200 pages

Publisher: Pragmatic Bookshelf; 1 edition (October 2, 2015)

Language: English

ISBN-10: 1680500708

ISBN-13: 978-1680500707

Product Dimensions: 7.5 x 0.6 x 9.2 inches

Shipping Weight: 1.1 pounds (View shipping rates and policies)

Average Customer Review: 4.7 out of 5 starsÂ Â See all reviewsÂ (3 customer reviews)

Best Sellers Rank: #756,860 in Books (See Top 100 in Books) #119 inÂ Books > Computers & Technology > Programming > Languages & Tools > Ruby #164 inÂ Books > Computers & Technology > Programming > Software Design, Testing & Engineering > Tools #578 inÂ Books > Computers & Technology > Databases & Big Data > Data Processing

Customer Reviews

Top Five Text Processing Tips by Rob Miller, author of Text Processing with Ruby Clean up your data first Data in the real world is messy. It almost always pays off to take some time to normalize different sources of data and to get them into the same format before you begin whatever actual processing you need to do. YouÃ¢Â Â™ll have less exceptions and special cases in your code, and itÃ¢Â Â™ll be a lot more resilient. Master regular expressions There are definitely some text processing problems that canÃ¢Â Â™t be solved with regular expressions, but not that many. While theyÃ¢Â Â™re not always the best or more readable option, knowing regular expressions well will get you out of many tight spots, and even more often than that will be the first step towards a more robust solution. Break your problem into discrete steps Almost all text processing tasks, no matter how complicated they seem on the face of it, are really a series of small transformations. Figuring out how to frame your problem in this way will make it easy to take a pipeline approach, where your text flows through a series of small, discrete steps, each of which transform the data in a particular way and then passes it on. Such programs are both easier to reason about and easier to modify and extend.

View larger Figure out a strategy for missing data Data in the real world, as well as being messy, also frequently has gaps. Decide early on how youÃ¢Â Â™re going to cope with that Ã¢Â Â” how youÃ¢Â Â™ll represent the absence of particular fields or properties Ã¢Â Â” and youÃ¢Â Â™ll avoid messiness later on. Make the most of existing tools There are hundreds of command-line tools that exist solely to process textual data. Each of them is capable of performing a particular transformation, which means you donÃ¢Â Â™t need to reinvent the wheel. If you use existing tools for the parts of your problem that have already been solved, all that remains is to solve the unique problem that you have.

Extract Order from Chaos Ruby: Programming, Master's Handbook: A TRUE Beginner's Guide! Problem Solving, Code, Data Science, Data Structures & Algorithms (Code like a PRO in ... web design, tech, perl, ajax, swift, python,) Data Visualization Toolkit: Using JavaScript, Rails, and Postgres to Present Data and Geospatial Information (Addison-Wesley Professional Ruby Series) The Abyss Surrounds Us Outcast by Kirkman & Azaceta Volume 1: A Darkness Surrounds Him Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking Ruby's Tea for Two (Max and Ruby) Ruby's Cupcakes (Max and Ruby) Ruby's Rainbow (Max and Ruby) Max & Ruby's Storybook Treasury (Max and Ruby) Ruby's Falling Leaves (Max and Ruby) Practical Object-Oriented Design in Ruby: An Agile Primer (Addison-Wesley Professional Ruby) Ruby on Rails Tutorial: Learn Web Development with Rails (3rd Edition) (Addison-Wesley Professional Ruby) Eloquent Ruby (Addison-Wesley Professional Ruby)