

This is a **low resolution, black and white** version of the article you downloaded. To download the whole of Free Software Magazine in *high* resolution and color, please subscribe!

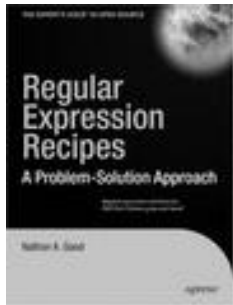
Subscriptions are free, and every subscriber receives our fantastic weekly newsletters — which are in fact fully edited articles about free software.

Please click here to subscribe:

<http://www.freesoftwaremagazine.com/subscribe>

# Regular Expression Recipes by Nathan A. Good

Martin C Brown



I'll admit right up front that I am something of a regular expression junkie. Years before I even knew such a system existed (before the days of the internet) I wrote my own regular expression system to handle the needs of a free-text database management package. Today, we are all familiar with regular expressions in Perl, sed, awk/gawk and even

in “user” applications like email and word processors.

Despite the utility of the regular expression library used in these systems, getting your regular expression right can be a major exercise. With the wrong expression you can match, or replace, the wrong text and in some applications that could have dire consequences.

Apress have released a new title, *Regular Expression Recipes*, written by Nathan A Good, that attempts to solve some of the riddles and complexities of the regular expression writing process by providing example regular expressions to be used in different situations and against a variety of problems.

---

For each problem, a full range of solutions is provided in different languages (perl, sed, etc.), using regular expressions and/or scripts relevant to that environment.

---

## The contents

Good has used an interesting approach to organizing the book. One of the major problems with regular expressions is that they are used in a variety of applications, sometimes with slight differences. Although the majority of modern applications use a derivation of the Perl regular expression library, some use their own, often out of a need for historical compatibility.

The book therefore starts off with a look at the various regular expression systems and a range of handy tables summarizing the differences between Perl, grep and Vim, which support the three main regular expression systems. Handier still, the book goes on to add details on how to use regular expressions in popular environments, such as Perl, Python, PHP, Vim, grep and sed.

This list of tools is relevant to the rest of the book as each example problem and regular expression solution is followed by two or more examples using the tools. In many cases these are full scripts or examples, although for some a simple fragment would be enough to get the idea. In addition, all of the scripts and regular expressions are explained and their operation detailed so that you understand why it works.

The bulk of the book is then split up into specific areas of issues that can be resolved with regular expressions, beginning with the more obvious area of basic text manipulation. Here we find the common—but not always straightforward—issues of finding words and lines and then finding and replacing text, for example capitalizing the first letter of a word, before moving on to the more complex issues of finding text in, or around, quotes.

Also handy in this section are a series of examples on replacing ASCII and non-ASCII characters with their equivalents. For example, replacing smart quotes with straight quotes or copyright and trademark symbols with a textual equivalent (i.e. <sup>TM</sup> with (tm)).

With the basics out of the way, the book moves on to more interesting topics. I won't list all the examples, or even all the different topic groups into which they are placed. I will say, however, that there isn't an example here that I thought was superfluous. Some highlights and personal favourites include, validating credit card numbers, extracting HTML attributes and a whole bunch on re-formatting code.

As stated earlier, for each problem, a full range of solutions is provided in different languages (perl, sed, etc.), using different regular expressions and scripts relevant to that environment. For example, one of the examples validates dates and example scripts and expressions are provided in Perl, PHP, grep and Vim.

## Who's this book for?

Regular expressions are one of the interesting parts of the computing world. Technically not a programming language, regular expressions are heavily used in programming. They are also common in a number of command line tools like grep and sed. All of this makes regular expressions, and by association this book, useful for just about any “power” user. Administrators will appreciate some of the tools for help in their scripts, log file parsing and when searching for information.

For programmers in particular, the book offers a wide range of examples and samples that can be used or adapted in applications.

Most of the samples can be used verbatim, others will probably benefit from direct modification according to your needs.

## Pros

The amazing range of problems and their solutions would be my first reason to recommend the purchase of this book. There is a regular expression example here for everybody. My second reason for recommendation would be the range of environments demonstrated through the examples. Use regular expressions, but not a Perl programmer? No problem, not only do you get the Perl sample, you get examples in an environment in which you might be more familiar, such as Vim or PHP. By covering each regular expression, and also details about why it works and examples for key environments like Perl or sed, the book becomes more than just a regular expression tool. This range means that the book is also an advanced scripting, programmers and administrators toolkit for performing a variety of tasks.

## Cons

I really couldn't find anything wrong with this book. Occasionally, I thought a sample in a particular environment was missing, but with such detailed information on the regular expression it really isn't that difficult to embed the expression into your own script. I'm really scraping the barrel here though; the book is without a doubt one of the best and I highly recommend it.

## In short

---

Title	Regular Expression Recipes
Author	Nathan A Good
Publisher	Apress
ISBN	159059441X
Year	2005
Pages	289
CD included	No
Mark	9

---

## Copyright information

© 2005 Martin C Brown

This article is made available under the "Attribution-NonCommercial-NoDerivs" Creative Commons License 2.0 available from <http://creativecommons.org/licenses/by-nc-nd/2.0/>.