

Im letzten Jahrzehnt ist die Beliebtheit regulärer Ausdrücke deutlich angestiegen. Heutzutage gibt es in allen verbreiteten Programmiersprachen mächtige Bibliotheken zur Verarbeitung regulärer Ausdrücke. Zum Teil bietet die Sprache sogar selbst die entsprechenden Möglichkeiten. Viele Entwickler nutzen diese Features, um den Anwendern ihrer Applikationen das Suchen und Filtern der Daten mithilfe regulärer Ausdrücke zu ermöglichen. Reguläre Ausdrücke sind überall.

Es gibt viele Bücher, die sich mit regulären Ausdrücken befassen. Die meisten machen ihren Job ganz gut – sie erklären die Syntax und enthalten ein paar Beispiele sowie eine Referenz. Aber es gibt keine Bücher, die Lösungen vorstellen. Lösungen, die auf regulären Ausdrücken basieren und für eine ganze Reihe von praktischen Problemen aus der realen Welt genutzt werden können. Bei solchen Problemen geht es vor allem um Fragen zu Texten auf einem Computer und um Internetanwendungen. Wir, Steve und Jan, haben uns dazu entschieden, diese Lücke mit diesem Buch zu füllen.

Wir wollten vor allem zeigen, wie Sie reguläre Ausdrücke in Situationen verwenden können, in denen weniger Erfahrene im Umgang mit regulären Ausdrücken sagen würden, das sei nicht möglich, oder in denen Softwarepuristen der Meinung sind, ein regulärer Ausdruck sei nicht das richtige Tool für diese Aufgabe. Da reguläre Ausdrücke heute überall zu finden sind, sind sie oft auch als Tool verfügbar, das von Endanwendern genutzt werden kann. Auch Programmierer können durch die Verwendung einiger weniger regulärer Ausdrücke viel Zeit sparen, etwa wenn sie Informationen suchen und verändern müssen. Die Alternative ist oft, Stunden oder Tage mit dem Umsetzen in prozeduralen Code zu verbringen oder eine Bibliothek von dritter Seite zu nutzen.

## Gefangen im Gewirr der verschiedenen Versionen

Vergleichbar mit anderen beliebten Dingen der IT-Branche, gibt es auch reguläre Ausdrücke in vielen unterschiedlichen Ausprägungen mit unterschiedlicher Kompatibilität. Das hat dazu geführt, dass es diverse Varianten eines regulären Ausdrucks gibt, die sich nicht immer gleich verhalten oder die teilweise gar nicht funktionieren.

Viele Bücher erwähnen, dass es unterschiedliche Varianten gibt, und führen auch einige der Unterschiede auf. Aber immer wieder lassen sie bestimmte Varianten unerwähnt – vor allem wenn in diesen Varianten Features fehlen -, statt auf alternative Lösungen und Workarounds hinzuweisen. Das ist frustrierend, wenn Sie mit unterschiedlichen Varianten regulärer Ausdrücke in den verschiedenen Anwendungen oder Programmiersprachen arbeiten müssen.

Saloppe Bemerkungen in der Literatur wie „jeder nutzt mittlerweile reguläre Ausdrücke im Perl-Stil“ bagatellisieren leider eine ganze Reihe von Inkompatibilitäten. Selbst Pakete im „Perl-Stil“ besitzen entscheidende

Unterschiede, und Perl entwickelt sich ja auch noch weiter. Solche oberflächlichen Äußerungen können für Programmierer trostlose Folgen haben und zum Beispiel dazu führen, dass sie eine halbe Stunde oder mehr damit verbringen, nutzlos im Debugger herumzustochern, statt die Details ihrer Implementierung für reguläre Ausdrücke zu kontrollieren. Selbst wenn sie herausfinden, dass ein Feature, auf dem sie aufbauen, nicht vorhanden ist, wissen sie nicht immer, wie sie stattdessen vorgehen können.

Dieses Buch ist das erste, das die am meisten verbreiteten und umfangreichsten Varianten regulärer Ausdrücke nebeneinander aufführt – und zwar durchgängig im ganzen Buch.

## Für wen dieses Buch gedacht ist

Sie sollten dieses Buch lesen, wenn Sie regelmäßig am Computer mit Text zu tun haben – egal ob Sie einen Stapel Dokumente durchsuchen, Text in einem Texteditor bearbeiten oder Software entwickeln, die Text durchsuchen oder verändern soll. Reguläre Ausdrücke sind für diese Aufgaben exzellente Hilfsmittel. Das Reguläre Ausdrücke Kochbuch erklärt Ihnen alles, was Sie über reguläre Ausdrücke wissen müssen. Sie brauchen kein Vorwissen, da wir selbst einfachste Aspekte regulärer Ausdrücke erklären werden.

Wenn Sie schon Erfahrung mit regulären Ausdrücken haben, werden Sie eine Menge Details kennenlernen, die in anderen Büchern oder Onlineartikeln häufig einfach übergangen werden. Sind Sie jemals über eine Regex gestolpert, die in einer Anwendung funktioniert hat, in einer anderen aber nicht, werden Sie die in diesem Buch detailliert und gleichwertig behandelten Beschreibungen zu sieben der verbreitetsten Varianten regulärer Ausdrücke sehr hilfreich finden. Wir haben das ganze Buch als Kochbuch aufgebaut, sodass Sie direkt zu den Themen springen können, die Sie interessieren. Wenn Sie dieses Buch von vorne bis hinten durchlesen, werden Sie am Ende Meister regulärer Ausdrücke sein.

Mit diesem Buch erfahren Sie alles, was Sie über reguläre Ausdrücke wissen müssen, und noch ein bisschen mehr – unabhängig davon, ob Sie Programmierer sind oder nicht. Wenn Sie reguläre Ausdrücke in einem Texteditor nutzen wollen, in einem Suchtool oder in irgendeiner Anwendung, die ein Eingabefeld „Regex“ enthält, können Sie dieses Buch auch ganz ohne Programmiererfahrung lesen. Die meisten Rezepte bieten Lösungen an, die allein auf einem oder mehreren regulären Ausdrücken basieren.

Die Programmierer unter Ihnen erhalten in Kapitel 3 alle notwendigen Informationen zum Implementieren regulärer Ausdrücke in ihrem Quellcode. Dieses Kapitel geht davon aus, dass Sie mit den grundlegenden Features der Programmiersprache Ihrer Wahl vertraut sind, aber Sie müssen keine Erfahrung mit dem Einsatz regulärer Ausdrücke in Ihrem Quellcode mitbringen.

## Behandelte Technologien

.NET, Java, JavaScript, PCRE, Perl, Python und Ruby kommen nicht ohne Grund auf dem Rückseitentext vor. Vielmehr stehen diese Begriffe für die sieben Varianten regulärer Ausdrücke, die in diesem Buch behandelt werden, wobei alle sieben gleichermaßen umfassend beschrieben werden. Insbesondere haben wir versucht, alle Uneinheitlichkeiten zu beschreiben, die wir in diesen verschiedenen Varianten finden konnten.

Das Kapitel zur Programmierung (Kapitel 3) enthält Code-Listings in C#, Java, JavaScript, PHP, Perl, Python, Ruby und VB.NET. Auch hier gibt es zu jedem Rezept Lösungen und Erläuterungen für alle acht Sprachen. Damit gibt es in diesem Kapitel zwar einige Wiederholungen, aber Sie können die Abhandlungen über Sprachen, an denen Sie nicht interessiert sind, gern überspringen, ohne etwas in der Sprache zu verpassen, die Sie selbst anwenden.

## Aufbau des Buchs

In den ersten drei Kapiteln dieses Buchs geht es um nützliche Tools und grundlegende Informationen, die eine Basis für die Verwendung regulärer Ausdrücke bilden. Jedes der folgenden Kapitel stellt dann eine Reihe von regulären Ausdrücken vor, die bestimmte Bereiche der Textbearbeitung behandeln.

**Kapitel 1, Einführung in reguläre Ausdrücke**, erläutert die Rolle regulärer Ausdrücke und präsentiert eine Reihe von Tools, die das Erlernen, Aufbauen und Debuggen erleichtern.

**Kapitel 2, Grundlagen regulärer Ausdrücke**, beschreibt alle Elemente und Features regulärer Ausdrücke zusammen mit wichtigen Hinweisen zu einer effektiven Nutzung.

**Kapitel 3, Mit regulären Ausdrücken programmieren**, stellt Coding-Techniken vor und enthält Codebeispiele für die Verwendung regulärer Ausdrücke in jeder der in diesem Buch behandelten Programmiersprachen.

**Kapitel 4, Validierung und Formatierung**, enthält Rezepte für den Umgang mit typischen Benutzereingaben, wie zum Beispiel Datumswerten, Telefonnummern und Postleitzahlen in den verschiedenen Staaten.

**Kapitel 5, Wörter, Zeilen und Sonderzeichen**, behandelt häufig auftretende Textbearbeitungsaufgaben, wie zum Beispiel das Testen von Zeilen auf die An- oder Abwesenheit bestimmter Wörter.

**Kapitel 6, Zahlen**, zeigt, wie man Integer-Werte, Gleitkommazahlen und viele andere Formate in diesem Bereich aufspürt.

**Kapitel 7, URLs, Pfade und Internetadressen**, zeigt Ihnen, wie Sie mit den Strings umgehen, die im Internet und in Windows-Systemen für das Auffinden von Inhalten genutzt werden.

**Kapitel 8, Markup und Datenaustausch**, dreht sich um das Bearbeiten von HTML, XML, Comma-Separated Values (CSV) und Konfigurationsdateien im INI-Stil.

## Die Codebeispiele verwenden

Dieses Buch ist dazu da, Ihnen bei Ihrer Arbeit zu helfen. Sie können den Code dieses Buchs in Ihren Programmen und Dokumentationen verwenden. Sie brauchen uns nicht um Erlaubnis zu fragen, solange Sie nicht einen beachtlichen Teil des Codes wiedergeben. Beispielsweise benötigen Sie keine Erlaubnis, um ein Programm zu schreiben, das einige Codeteile aus diesem Buch verwendet. Für den Verkauf oder die Verbreitung einer CD-ROM mit Beispielen aus O'Reilly-Büchern brauchen Sie auf jeden Fall unsere Erlaubnis. Die Beantwortung einer Frage durch das Zitieren dieses Buchs und seiner Codebeispiele benötigt wiederum keine Erlaubnis. Wenn Sie einen erheblichen Teil der Codebeispiele dieses Buchs in die Dokumentation Ihres Produkts einfügen, brauchen Sie eine Erlaubnis.

Wir freuen uns über einen Herkunftsnachweis, bestehen aber nicht darauf. Eine Referenz enthält i.d.R. Titel, Autor, Verlag und ISBN, zum Beispiel: „Reguläre Ausdrücke Kochbuch von Jan Goyvaerts & Steven Levithan, Copyright 2010, O'Reilly Verlag, ISBN 978-3-89721-957-1.“

Wenn Sie denken, Ihre Verwendung unserer Codebeispiele könnte den angemessenen Gebrauch oder die hier erteilte Erlaubnis überschreiten, nehmen Sie einfach mit uns über [kommentar@oreilly.de](mailto:kommentar@oreilly.de) Kontakt auf.

## Danksagung

Wir danken Andy Oram, unserem Lektor bei O'Reilly Media, Inc., für seine Begleitung bei diesem Projekt vom Anfang bis zum Ende. Ebenso danken wir Jeffrey Friedl, Zak Greant, Nikolaj Lindberg und Ian Morse für ihre sorgfältigen fachlichen Korrekturen, durch die dieses Buch umfassender und genauer wurde.

## Über die Autoren

Das Reguläre Ausdrücke Kochbuch wurde von Jan Goyvaerts und Steven Levichan geschrieben, zwei der weltweit anerkanntesten Experten auf dem Gebiet

der regulären Ausdrücke.

Jan Goyvaerts betreibt das Unternehmen Just Great Software, aus dem einige der beliebtesten Software-Tools für reguläre Ausdrücke stammen. Dazu gehören RegexBuddy, der weltweit einzige Regex-Editor, der die Besonderheiten von 15 Regex-Varianten abbildet, und PowerGREP, das umfassendste Grep-Tool für Windows.

Steven Levithan ist ein führender Experte für reguläre Ausdrücke im JavaScript-Umfeld und betreibt einen Blog zum Thema reguläre Ausdrücke: <http://blog.stevenlevithan.com>. Seit Jahren betrachtet er es als sein Hobby, sein Wissen über die verschiedenen Varianten der regulären Ausdrücke und die dazugehörigen Libraries auszubauen.