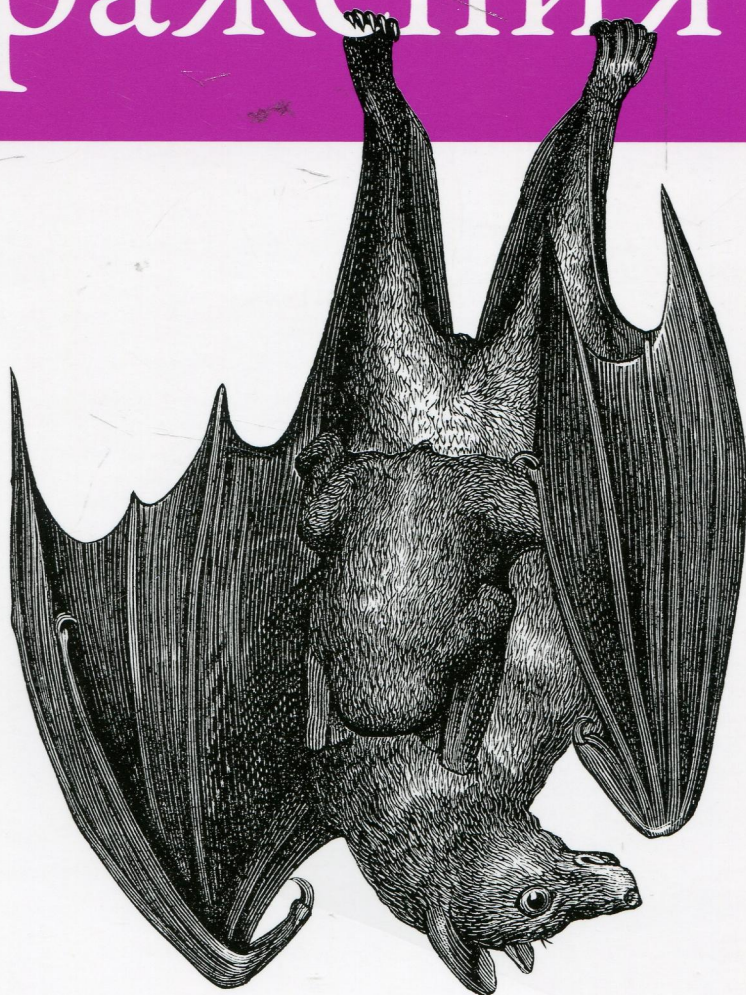


Лаконично, доходчиво, пошагово

Основы

Регулярные выражения



O'REILLY®

Майкл Фицджеральд

Регулярные выражения

Основы

Майкл Фицджеральд



Москва • Санкт-Петербург • Киев
2016

ББК 32.973.26-018.2.75

Ф66

УДК 681.3.07

Издательский дом “Вильямс”

Главный редактор С.Н. Тригуб

Зав. редакцией В.Р. Гинзбург

Перевод с английского и редакция канд. хим. наук А.Г. Гузиковича

По общим вопросам обращайтесь в Издательский дом “Вильямс” по адресу:

info@williamspublishing.com, <http://www.williamspublishing.com>

Фицджеральд, Майкл.

Ф66 Регулярные выражения: основы. : Пер. с англ. — М. : ООО “И.Д. Вильямс”, 2016. — 144 с. : ил. — Парал. тит. англ.

ISBN 978-5-8459-1953-3 (рус.)

ББК 32.973.26-018.2.75

Все названия программных продуктов являются зарегистрированными торговыми марками соответствующих фирм.

Никакая часть настоящего издания ни в каких целях не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами, будь то электронные или механические, включая фотокопирование и запись на магнитный носитель, если на это нет письменного разрешения издательства O'Reilly Media, Inc.

Authorized Russian translation of the English edition of *Introducing Regular Expressions* (ISBN 978-1-449-39268-0). Copyright © 2012 Michael Fitzgerald.

This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without the prior written permission of the copyright owner and the Publisher.

Книга отпечатана согласно договору с ООО “Дальрегионсервис”.

Научно-популярное издание

Майкл Фицджеральд

Регулярные выражения: основы

Литературный редактор *И.А. Попова*

Верстка *О.В. Мишутина*

Художественный редактор *Е.П. Дынник*

Корректор *Л.А. Гордиенко*

Подписано в печать 11.04.2016. Формат 70x100/16

Гарнитура Times

Усл. печ. л. 11,61. Уч.-изд. л. 7,7

Доп. тираж 200 экз. Заказ № 2211

Отпечатано в АО «Первая Образцовая типография»

филиал «Чеховский Печатный Двор»

142300, Московская область, г. Чехов, ул. Полиграфистов, д.1

Сайт: www.chpd.ru, E-mail: sales@chpd.ru, тел. 8(499)270-73-59

ООО “И. Д. Вильямс”, 127055, г. Москва, ул. Лесная, д. 43, стр. 1

ISBN 978-5-8459-1953-3 (рус.)

ISBN 978-1-449-39268-0 (англ.)

© 2015 Издательский дом “Вильямс”

© 2012 Michael Fitzgerald

Оглавление

Введение	11
Глава 1. Что такое регулярные выражения	15
Глава 2. Сопоставление с простыми шаблонами	27
Глава 3. Границы	43
Глава 4. Альтернативы, группы и обратные ссылки	55
Глава 5. Символьные классы	67
Глава 6. Сопоставление с символами Unicode и другими символами	75
Глава 7. Квантификаторы	87
Глава 8. Группы проверки	95
Глава 9. Разметка документа тегами HTML5	101
Глава 10. Конец начала	115
Приложение А. Справочник по регулярным выражениям	121
Глоссарий	136
Предметный указатель	144

Содержание

Об авторе	9
Об изображении на обложке	9
Введение	11
Для кого предназначена эта книга	12
Что необходимо для работы с книгой	13
Соглашения, принятые в книге	13
Использование кода примеров	14
Ждем ваших отзывов!	14
Глава 1. Что такое регулярные выражения	15
Приложение Regexpal	16
Соответствие телефонному номеру	17
Задание соответствия цифрам с помощью символьных классов	18
Использование символьных сокращений	18
Соответствие произвольному одиночному символу	19
Захватывающие группы и обратные ссылки	19
Использование квантификаторов	20
Использование литеральных круглых скобок	21
Приложения со встроенным механизмом регулярных выражений	23
О чем вы узнали в главе 1	25
На заметку	25
Глава 2. Сопоставление с простыми шаблонами	27
Соответствие литеральному тексту	29
Соответствие цифрам	29
Соответствие нецифровым символам	31
Соответствие словарным и несловарным символам	32
Соответствие пробелам	34
Соответствие произвольному символу	35
Разметка текста тегами	38
Использование редактора <i>sed</i> для разметки текста	38
Использование Perl для разметки текста	39
О чем вы узнали в главе 2	41
На заметку	41
Глава 3. Границы	43
Начало и конец строки	43
Позиции, являющиеся и не являющиеся границами слов	45
Другие якорные привязки	47
Задание группы символов как литералов	48
Добавление тегов	49
Добавление тегов с помощью <i>sed</i>	50
Добавление тегов с помощью Perl	51

О чем вы узнали в главе 3	52
На заметку	52
Глава 4. Альтернативы, группы и обратные ссылки	55
Чередование	55
Подшаблоны	59
Захватывающие группы и обратные ссылки	60
Именованные группы	62
Незахватывающие группы	64
Атомарные группы	64
О чем вы узнали в главе 4	65
На заметку	65
Глава 5. Символьные классы	67
Инвертированные символьные классы	69
Объединение и разность	70
Символьные классы POSIX	71
О чем вы узнали в главе 5	73
На заметку	73
Глава 6. Сопоставление с символами Unicode и другими символами	75
Сопоставление с символами Unicode	76
Использование редактора <i>vim</i>	78
Поиск соответствий восьмеричным кодам символов	79
Поиск соответствий свойствам символов Unicode	79
Поиск соответствий управляющим символам	83
О чем вы узнали в главе 6	85
На заметку	85
Глава 7. Квантификаторы	87
Жадный, ленивый и сверхжадный поиск	88
Сопоставление с использованием квантификаторов *, + и ?	88
Соответствие заданному количеству повторений символа	90
Ленивые квантификаторы	91
Сверхжадные квантификаторы	92
О чем вы узнали в главе 7	93
На заметку	94
Глава 8. Группы проверки	95
Положительная опережающая проверка	95
Отрицательная опережающая проверка	98
Положительная ретроспективная проверка	99
Отрицательная ретроспективная проверка	99
О чем вы узнали в главе 8	100
На заметку	100

Глава 9. Разметка документа тегами HTML5	101
Сопоставление с тегами	101
Преобразование простого текста с помощью редактора <i>sed</i>	103
Замена текста с помощью редактора <i>sed</i>	103
Обработка римских цифр в редакторе <i>sed</i>	104
Обработка отдельного абзаца в редакторе <i>sed</i>	105
Обработка строк поэмы в редакторе <i>sed</i>	105
Добавление тегов	106
Использование командного файла в редакторе <i>sed</i>	107
Преобразование простого текста с помощью Perl	108
Обработка римских цифр с помощью Perl	110
Обработка отдельного абзаца с помощью Perl	110
Обработка строк поэмы с помощью Perl	110
Использование командного файла в Perl	111
О чем вы узнали в главе 9	112
На заметку	113
Глава 10. Конец начала	115
Что дальше	116
Инструменты, реализации и библиотеки	117
Perl	117
PCRE	117
Ruby (Oniguruma)	118
Python	118
RE2	118
Сопоставление с телефонными номерами в формате, принятом в США и Канаде	119
Сопоставление с адресами электронной почты	120
О чем вы узнали в главе 10	120
Приложение А. Справочник по регулярным выражениям	121
Регулярные выражения в QED	121
Метасимволы	122
Специальные символы	123
Пробельные символы	124
Пробельные символы Unicode	124
Управляющие символы	125
Свойства символов	127
Имена шрифтов в свойствах символов	128
Символьные классы POSIX	129
Опции и модификаторы	130
Таблица кодов ASCII и представление ASCII-символов в регулярных выражениях	130
На заметку	135
Глоссарий	136
Предметный указатель	144