

# Оглавление

<b>Предисловие</b> .....	<b>16</b>
Почему я написал эту книгу .....	16
Для кого написана эта книга .....	17
Как читать эту книгу .....	18
Структура книги .....	18
Подробное описание .....	20
Информация по инструментам .....	20
Условные обозначения .....	21
Упражнения .....	22
Ссылки, программный код, ошибки и контакты .....	23
Личные комментарии и благодарности .....	23
От издательства .....	24
<b>Глава 1. Знакомство с регулярными выражениями</b> .....	<b>25</b>
Решение реальных задач .....	26
Регулярные выражения как язык .....	28
Аналогия с файловыми шаблонами .....	28
Аналогия с языками .....	29
Регулярные выражения как особый склад ума .....	30
Для читателей, имеющих опыт работы с регулярными выражениями .....	30
Поиск в текстовых файлах: egrep .....	31
Метасимволы egrep .....	32
Начало и конец строки .....	33
Символьные классы .....	33
Один произвольный символ .....	36
Выбор .....	37
Игнорирование различий в регистре символов .....	40
Границы слов .....	40

В двух словах . . . . .	42
Необязательные элементы . . . . .	43
Другие квантификаторы: повторение . . . . .	44
Круглые скобки и обратные ссылки . . . . .	47
Экранирование . . . . .	49
Новые горизонты . . . . .	50
Языковая диверсификация . . . . .	50
Смысл регулярного выражения . . . . .	50
Дополнительные примеры . . . . .	51
Терминология регулярных выражений . . . . .	54
Пути к совершенствованию . . . . .	58
Заключение . . . . .	60
Личные заметки . . . . .	62
<b>Глава 2. Дополнительные примеры . . . . .</b>	<b>63</b>
О примерах . . . . .	64
Краткий курс Perl . . . . .	65
Поиск по регулярному выражению . . . . .	66
Переходим к реальным примерам . . . . .	69
Побочные эффекты успешных совпадений . . . . .	69
Взаимодействие регулярных выражений с логикой программы . . . . .	72
Лирическое отступление . . . . .	79
Модификация текста с использованием регулярных выражений . . . . .	80
Пример: письмо на стандартном бланке . . . . .	81
Пример: обработка биржевых котировок . . . . .	82
Автоматизация редактирования . . . . .	83
Маленькая почтовая утилита . . . . .	84
Разделение разрядов числа запятыми . . . . .	91
Преобразование текста в HTML . . . . .	101
Задача с повторяющимися словами . . . . .	112
<b>Глава 3. Регулярные выражения: возможности и диалекты . . . . .</b>	<b>118</b>
История регулярных выражений . . . . .	120
Происхождение регулярных выражений . . . . .	120
На первый взгляд . . . . .	128
Основные операции с регулярными выражениями . . . . .	130
Интегрированный интерфейс . . . . .	131

Процедурный и объектно-ориентированный интерфейс . . . . .	132
Поиск с заменой . . . . .	136
Поиск и замена в других языках . . . . .	138
Итоги . . . . .	140
Строки, кодировки и режимы . . . . .	140
Строки как регулярные выражения . . . . .	140
Проблемы кодировки символов . . . . .	145
Юникод . . . . .	146
Режимы обработки регулярных выражений и поиска совпадений . . . . .	150
Стандартные метасимволы и возможности . . . . .	154
Представления символов . . . . .	156
Символьные классы и их аналоги . . . . .	161
Якорные метасимволы и другие проверки с нулевой длиной совпадения . . . . .	175
Комментарии и модификаторы режимов . . . . .	182
Группировка, сохранение, условные и управляющие конструкции . . . . .	184
Путеводитель по серьезным главам . . . . .	191

**Глава 4. Механика обработки регулярных выражений . . . . . 192**

Запустить двигатели! . . . . .	192
Два вида двигателей . . . . .	192
Новые стандарты . . . . .	193
Типы механизмов регулярных выражений . . . . .	194
С позиций избыточности . . . . .	195
Определение типа механизма . . . . .	196
Основы поиска совпадений . . . . .	197
О примерах . . . . .	197
Правило 1: более раннее совпадение выигрывает . . . . .	198
Компоненты и части двигателя . . . . .	199
Правило 2: квантификаторы работают максимально . . . . .	201
Механизмы регулярных выражений . . . . .	204
НКА: механизм, управляемый регулярным выражением . . . . .	204
ДКА: механизм, управляемый текстом . . . . .	206
Сравнение двух механизмов . . . . .	207
Возврат . . . . .	209
Крошечная аналогия . . . . .	209
Два важных замечания . . . . .	210

Сохраненные состояния	211
Возврат и максимализм	214
Подробнее о максимализме и о возврате	217
Проблемы максимализма	217
Многосимвольные «кавычки»	218
Минимальные квантификаторы	219
Максимальные и минимальные конструкции всегда выбирают совпадение	221
О сущности максимализма, минимализма и возврата	222
Захватывающие квантификаторы и атомарная группировка	223
Захватывающие квантификаторы <code>?</code> , <code>*</code> , <code>++</code> и <code>{max, min}+</code>	227
Возврат при позиционной проверке	228
Максимальна ли конструкция выбора?	229
Использование упорядоченного выбора	230
НКА, ДКА и POSIX	233
«Самое длинное совпадение, ближе к левому краю»	233
POSIX и правило «самого длинного совпадения, ближнего к левому краю»	234
Скорость и эффективность	235
Сравнение ДКА и НКА	237
Итоги	240

## **Глава 5. Практические приемы построения регулярных выражений . . . . . 242**

Балансировка регулярных выражений	243
Несколько коротких примеров	243
Снова о строках продолжения	243
Поиск IP-адреса	244
Работа с именами файлов	247
Поиск парных скобок	251
Исключение нежелательных совпадений	253
Поиск текста в ограничителях	254
Данные и предположения	258
Удаление пропусков в начале и конце строки	258
Работа с HTML	260
Поиск тегов HTML	260
Поиск ссылок HTML	261

Анализ HTTP URL . . . . .	263
Проверка имени хоста . . . . .	265
Поиск URL на практике . . . . .	267
Нетривиальные примеры . . . . .	271
Синхронизация . . . . .	271
Разбор данных, разделенных запятыми . . . . .	275

## **Глава 6. Построение эффективных регулярных выражений . . . . . 283**

Убедительный пример . . . . .	284
Простое изменение — начинаем с более вероятного случая . . . . .	285
Эффективность и правильность . . . . .	286
Следующий шаг — локализация максимального поиска . . . . .	286
Возвращение к реальности . . . . .	289
Возврат с глобальной точки зрения . . . . .	291
POSIX НКА — работа продолжается . . . . .	292
Работа механизма при отсутствии совпадения . . . . .	293
Уточнение . . . . .	294
Конструкция выбора может дорого обойтись . . . . .	295
Хронометраж . . . . .	295
Зависимость результатов хронометража от данных . . . . .	298
Хронометраж в языке PHP . . . . .	298
Хронометраж в языке Java . . . . .	299
Хронометраж в языке VB.NET . . . . .	301
Хронометраж в языке Ruby . . . . .	302
Хронометраж в языке Python . . . . .	303
Хронометраж в языке Tcl . . . . .	304
Стандартные оптимизации . . . . .	305
Ничто не дается бесплатно . . . . .	305
Универсальных истин не бывает . . . . .	306
Механика применения регулярных выражений . . . . .	306
Предварительные оптимизации . . . . .	308
Оптимизации при смещении текущей позиции . . . . .	312
Оптимизации на уровне регулярных выражений . . . . .	314
Приемы построения быстрых выражений . . . . .	320
Приемы, основанные на здравом смысле . . . . .	322
Выделение литерального текста . . . . .	323
Выделение якорей . . . . .	324

Выбор между минимальными и максимальными квантификаторами . . . . .	325
Разделение регулярных выражений . . . . .	325
Имитация исключения по первому символу . . . . .	327
Использование атомарной группировки и захватывающих квантификаторов . . . . .	328
Руководство процессом поиска . . . . .	329
Раскрутка цикла . . . . .	330
Метод 1: построение регулярного выражения по результатам тестов . . . . .	331
Общий шаблон «раскрутки цикла» . . . . .	333
Метод 2: структурный анализ . . . . .	337
Метод 3: имена хостов Интернета . . . . .	337
Замечания . . . . .	339
Применение атомарной группировки и захватывающих квантификаторов . . . . .	339
Примеры раскрутки цикла . . . . .	341
Раскрутка комментариев C . . . . .	343
Исключение случайных совпадений . . . . .	349
Управление поиском совпадения . . . . .	350
Управление поиском = скорость . . . . .	352
Свертка . . . . .	354
Вывод: думайте! . . . . .	355

## **Глава 7. Perl . . . . . 356**

Регулярные выражения как компонент языка . . . . .	358
Самая сильная сторона Perl . . . . .	360
Самая слабая сторона Perl . . . . .	360
Диалект регулярных выражений Perl . . . . .	360
Регулярные выражения — операнды и литералы . . . . .	363
Порядок обработки литералов регулярных выражений . . . . .	367
Модификаторы регулярных выражений . . . . .	368
Perl'измы из области регулярных выражений . . . . .	369
Контекст выражения . . . . .	370
Динамическая видимость и последствия совпадения регулярных выражений . . . . .	371
Специальные переменные, изменяемые при поиске . . . . .	376
Оператор qr/.../ и объекты регулярных выражений . . . . .	381
Построение и использование объектов регулярных выражений . . . . .	382
Просмотр содержимого объектов регулярных выражений . . . . .	384
Объекты регулярных выражений и повышение эффективности . . . . .	385

Оператор поиска	385
Операнд регулярное выражение	386
Операнд целевой текст	387
Варианты использования оператора поиска	389
Интерактивный поиск — скалярный контекст с модификатором /g	392
Внешние связи оператора поиска	398
Оператор подстановки	400
Операнд-замена	400
Модификатор /e	401
Контекст и возвращаемое значение	403
Оператор разбиения	403
Простейшее разбиение	404
Возвращение пустых элементов	406
Специальные значения первого операнда split	408
Сохраняющие круглые скобки в первом операнде split	409
Специфические возможности Perl	410
Применение динамических регулярных выражений для поиска вложенных конструкций	412
Встроенный код	415
Ключевое слово local во встроенном коде	421
Встроенный код и переменные my	424
Поиск вложенных конструкций	426
Перегрузка литералов регулярных выражений	428
Ограничения перегрузки литералов регулярных выражений	431
Имитация именованного сохранения	433
Проблемы эффективности в Perl	435
У каждой задачи есть несколько решений	436
Компиляция регулярных выражений, модификатор /o, qr/.../ и эффективность	438
Предварительное копирование	444
Функция study	449
Хронометраж	450
Отладочная информация регулярных выражений	451
Последний комментарий	454
<b>Глава 8. Java</b>	<b>456</b>
Диалект регулярных выражений	457
Поддержка конструкций \p{...} и \P{...} в Java	460

Завершители строк Юникода . . . . .	462
Использование пакета <code>java.util.regex</code> . . . . .	463
Метод <code>Pattern.compile()</code> . . . . .	464
Метод <code>Pattern.matcher()</code> . . . . .	465
Объект <code>Matcher</code> . . . . .	466
Применение регулярного выражения . . . . .	468
Получение информации о результатах . . . . .	470
Простой поиск с заменой . . . . .	472
Расширенный поиск с заменой . . . . .	475
Поиск с заменой по месту . . . . .	477
Область в объекте <code>Matcher</code> . . . . .	479
Объединение методов в конвейер . . . . .	486
Методы для построения сканеров . . . . .	487
Другие методы <code>Matcher</code> . . . . .	490
Другие методы <code>Pattern</code> . . . . .	492
Метод <code>split</code> класса <code>Pattern</code> с одним аргументом . . . . .	493
Метод <code>split</code> класса <code>Pattern</code> с двумя аргументами . . . . .	494
Дополнительные примеры . . . . .	496
Добавление атрибутов <code>WIDTH</code> и <code>HEIGHT</code> в теги <code>&lt;img&gt;</code> . . . . .	496
Проверка корректности HTML-кода с использованием нескольких регулярных выражений на один объект <code>Matcher</code> . . . . .	498
Разбор данных CSV . . . . .	499
Различия между версиями Java . . . . .	502
Различия между 1.4.2 и 1.5.0 . . . . .	502
Различия между 1.5.0 и 1.6.0 . . . . .	504
<b>Глава 9. .NET . . . . .</b>	<b>505</b>
Диалект регулярных выражений <code>.NET</code> . . . . .	506
Замечания по поводу диалекта <code>.NET</code> . . . . .	510
Использование регулярных выражений в <code>.NET</code> . . . . .	515
Основные принципы работы с регулярными выражениями . . . . .	515
Общие сведения о пакете . . . . .	518
Краткая сводка основных объектов . . . . .	519
Основные объекты . . . . .	521
Создание объектов <code>Regex</code> . . . . .	522
Использование объектов <code>Regex</code> . . . . .	525
Использование объектов <code>Match</code> . . . . .	534
Использование объектов <code>Group</code> . . . . .	535



Статические вспомогательные функции	536
Кэширование регулярных выражений	537
Дополнительные функции	538
Нетривиальные возможности .NET	540
Сборки регулярных выражений	540
Поиск вложенных конструкций	541
Объект Capture	544

## **Глава 10. PHP** ..... **546**

Диалект регулярных выражений PHP	548
Функциональный интерфейс механизма preg	551
Аргумент «шаблон»	552
Функции preg	559
preg_match	559
preg_match_all	565
preg_replace	571
preg_replace_callback	577
preg_split	580
preg_grep	586
preg_quote	587
«Недостающие» функции preg	588
preg_regex_to_pattern	588
Проверка синтаксиса неизвестного шаблона	591
Проверка синтаксиса неизвестного регулярного выражения	593
Рекурсивные регулярные выражения	593
Поиск совпадений с вложенными круглыми скобками	593
Никаких возвратов в рекурсии	596
Совпадение с парой вложенных скобок	596
Вопросы эффективности в PHP	597
Модификатор шаблона S: «Study»	597
Расширенные примеры	600
Разбор данных в формате CVS в PHP	600
Проверка тегированных данных на корректность вложенных конструкций	601