

# Введение

Эта книга научит вас работать с регулярными выражениями на конкретных примерах. Цель книги — максимально упростить освоение регулярных выражений. Фактически каждое понятие обсуждается с приведением соответствующих примеров, которые читателю будет легко повторить и проверить.

Регулярные выражения упрощают поиск определенных образцов текста. Точнее говоря, они представляют собой текстовые строки, описывающие на специальном языке искомые шаблонные комбинации символов в наборах текстовых строк, в большинстве случаев — строк, хранящихся в документах или файлах.

Формальная теория регулярных выражений была впервые изложена математиком Стивенем Клином в его книге *Introduction to Metamathematics* (New York, Van Nostrand), опубликованной в 1952 году, однако ее основные концепции были разработаны еще в начале 1940-х годов. Широкую популярность среди компьютерных специалистов регулярные выражения приобрели в начале 1970-х годов после выхода операционной системы Unix (детища Брайана Кернигана, Денниса Ритчи, Кена Томпсона и других сотрудников корпорации AT&T Bell Laboratories), включающей такие утилиты, как *sed* и *grep*.

Насколько мне известно, одним из первых компьютерных приложений, в которых начали использоваться регулярные выражения, был текстовый редактор QED (сокр. от “Quick Editor” — быстрый редактор). Код этого редактора был написан для системы распределения времени Berkley Time-Sharing System, выполнявшейся на компьютере SDS 940 компании Scientific Data Systems. Версия QED, задокументированная в 1970 году, — это переписанный Кеном Томпсоном вариант существовавшего в то время редактора для системы Compatible Time-Sharing System, который был разработан сотрудниками вычислительного центра MIT и включал одну из ранних, если не самую первую, практических реализаций регулярных выражений для вычислительных целей. (Возможности регулярных выражений, предлагаемые в редакторе QED, описаны в табл. А.1 приложения.)

Для демонстрации примеров в книге применяются различные средства, и я надеюсь, что большинство из них окажутся для вас полезными и их использование не вызовет трудностей, однако некоторые средства могут быть недоступны для пользователей Windows. Конечно, вы сможете пропускать описания примеров, которые вам не удастся воспроизвести в силу отсутствия того или иного средства. Однако я считаю, что каждый, кто всерьез задумывается о карьере компьютерного специалиста, должен обязательно ознакомиться с методами обработки регулярных выражений в Unix-подобных средах. Я работаю с подобными системами вот уже 25 лет и все равно каждый день нахожу для себя что-то новое.

“Бедняги, не знакомые с UNIX, обречены заново изобретать велосипед”, — Генри Спенсер

К некоторым из представленных в книге инструментов возможен доступ в Интернете с помощью браузера, что будет наиболее удобно для большинства читателей. Часть инструментов требует использования командной строки, тогда как другие инструменты доступны в виде настольных приложений. В случае отсутствия у вас нужного инструментария его можно легко загрузить из Интернета. Большинство инструментальных средств, о которых пойдет речь, бесплатны или стоят совсем недорого.

В этой книге я стараюсь описывать регулярные выражения простым и понятным языком. Специальная терминология употребляется весьма экономно и только в самой необходимой степени. Я придерживаюсь именно такого подхода, поскольку многолетний опыт научил меня, что чрезмерно насыщенный терминами текст нередко препятствует пониманию сути. Это соответствует принципу, положенному в основу книги: много полезного можно делать даже тогда, когда еще не до конца вник в суть проблемы.

Существует множество различных реализаций регулярных выражений. В частности, вы увидите, что они используются в таких инструментах командной строки Unix, как *vi* (*vim*), *grep* или *sed*. Регулярные выражения встроены в такие языки программирования, как Perl (а разве могло быть иначе?), Java, JavaScript, C#, Ruby и многие другие, а также в декларативные языки наподобие XSLT 2.0. Список реализаций может быть продолжен такими настольными приложениями, как Notepad++, Oxygen или TextMate.

Большинство из указанных реализаций регулярных выражений в чем-то сходны, а в чем-то различаются. Я не могу подробно обсудить все отличия в столь маленькой книге, но о многих расскажу. Любые попытки задокументировать *все* различия между *всеми* реализациями наверняка привели бы меня в больницу. Поэтому углубляться во все подробности такого рода я не буду. Данная книга предназначена лишь для ознакомления читателя с регулярными выражениями, и эта цель будет достигнута.

## Для кого предназначена эта книга

Предполагаемые читатели книги — люди, которые за всю свою жизнь не написали еще ни одного регулярного выражения. Если вы новичок в этой области или в программировании вообще, то книга будет для вас хорошим вводным курсом. Другими словами, она предназначена для тех читателей, которые слышали кое-что о регулярных выражениях и заинтересовались ими, но пока еще не до конца понимают, что они собой представляют. Если вы относитесь к данной категории, то эта книга как раз для вас.

В целом, рассматривая свойства регулярных выражений, я буду придерживаться принципа “от простого к сложному”. Иными словами, новые сведения будут преподноситься постепенно, небольшими порциями.

Если вы достаточно хорошо знакомы с регулярными выражениями и уверенно ими пользуетесь, вам лучше обратиться к другим книгам. Эта книга предназначена для новичков, которые нуждаются в том, чтобы их буквально вели за руку. Если вам уже приходилось сталкиваться с регулярными выражениями, но ваша практика работы с ними довольно ограничена, то эта книга будет для вас полезной. Однако темпы рассмотрения материала могут показаться вам не столь быстрыми, как хотелось бы.

Могу порекомендовать несколько учебных пособий, которые имеет смысл изучить после прочтения данной книги. В книге Джеффри Фридла *Регулярные выражения, 3-е издание* (Символ-Плюс, 2008 г.) регулярные выражения рассмотрены гораздо более подробно. Кроме того, можете прочитать книгу Яна Гойвертса и Стивена Левитана *Регулярные*

выражения. *Сборник рецептов, 2-е издание* (Символ-Плюс, 2015 г.). Ян Гойвертс — создатель RegexBuddy, мощного настольного приложения для работы с регулярными выражениями (<http://www.regexbuddy.com/>), тогда как Стивен Левитан создал RegexPal, онлайн-процессор регулярных выражений (<http://www.regexpal.com>), который будет использован в первой главе.

## Что необходимо для работы с книгой

Чтобы чтение книги принесло вам максимальную пользу, в вашем распоряжении должны быть некоторые средства Unix (Linux), доступ к которым на компьютерах Mac обеспечивается установкой операционной системы Darwin (разновидность BSD), а на компьютерах Windows — установкой приложения Cygwin, в дистрибутив которого включены многие инструменты GNU (см. <http://www.cygwin.com> и <http://www.gnu.org>).

Книга изобилует многочисленными примерами. Можете просто просматривать их, но, для того чтобы действительно чему-то научиться, лучше самостоятельно выполнить как можно больше примеров, поскольку, по моему глубокому убеждению, важнейший фактор обучения — закрепление получаемых знаний на практике. Я представлю вам веб-приложения, упрощающие освоение регулярных выражений за счет цветового выделения найденных совпадений, а также познакомлю с инструментальными “рабочими лошадками” из мира Unix и настольными приложениями для анализа регулярных выражений и их использования в контекстном поиске.

Примеры из книги можно найти на сайте Github по следующему адресу:

```
https://github.com/michaeljamesfitzgerald/  
Introducing-Regular-Expressions
```

Кроме того, полный архив всех примеров и тестовых файлов, используемых в книге, доступен для загрузки по следующим адресам:

```
http://examples.oreilly.com/0636920012337/examples.zip  
http://www.williamsublishing.com/Books/978-5-8459-1953-3.html
```

Будет лучше, если вы сначала создадите рабочий каталог или папку на своем компьютере и загрузите туда эти файлы, а затем уже приступите к чтению книги.

## Соглашения, принятые в книге

В книге использованы следующие типографские соглашения.

- *Курсивом* выделяются новые термины, URL-адреса, адреса электронной почты, имена и расширения имен файлов и т.п.
- Моноширинный шрифт используется в листингах программ, а также в основном тексте для представления регулярных выражений, содержимого командной строки и других подобных элементов.
- Этой пиктограммой обозначены разделы текста, содержащие советы, рекомендации, а также замечания общего характера.



## Использование кода примеров

Эта книга была написана для того, чтобы облегчить вам работу. Вообще говоря, вы можете свободно использовать приведенный в книге код в своих программах и документации. Получения какого-либо специального разрешения от нас, если только речь не идет о значительных объемах кода, не требуется. Например, использование в вашей программе нескольких фрагментов кода, взятых из книги, не требует разрешения. Однако продажа или распространение компакт-диска, содержащего примеры из книг, выпущенных издательством O'Reilly, без предварительного получения разрешения запрещена. Цитирование данной книги и использование кода примеров в ответах на вопросы не требует разрешения. Вместе с тем, если вы включаете в документацию своего продукта значительные объемы кода из приведенных в книге примеров, то получение соответствующего разрешения является обязательным условием.

## Ждем ваших отзывов!

Вы, читатель этой книги, и есть главный ее критик. Мы ценим ваше мнение и хотим знать, что было сделано нами правильно, что можно было сделать лучше и что еще вы хотели бы увидеть изданным нами. Нам интересны любые ваши замечания в наш адрес.

Мы ждем ваших комментариев и надеемся на них. Вы можете прислать нам бумажное или электронное письмо либо просто посетить наш сайт и оставить свои замечания там. Одним словом, любым удобным для вас способом дайте нам знать, нравится ли вам эта книга, а также выскажите свое мнение о том, как сделать наши книги более интересными для вас.

Отправляя письмо или сообщение, не забудьте указать название книги и ее авторов, а также свой обратный адрес. Мы внимательно ознакомимся с вашим мнением и обязательно учтем его при отборе и подготовке к изданию новых книг.

Наши электронные адреса:

E-mail: [info@williamspublishing.com](mailto:info@williamspublishing.com)

WWW: [www.williamspublishing.com](http://www.williamspublishing.com)

Наши почтовые адреса:

в России: 127055, Москва, ул. Лесная, д. 43, стр. 1

в Украине: 03150, Киев, а/я 152