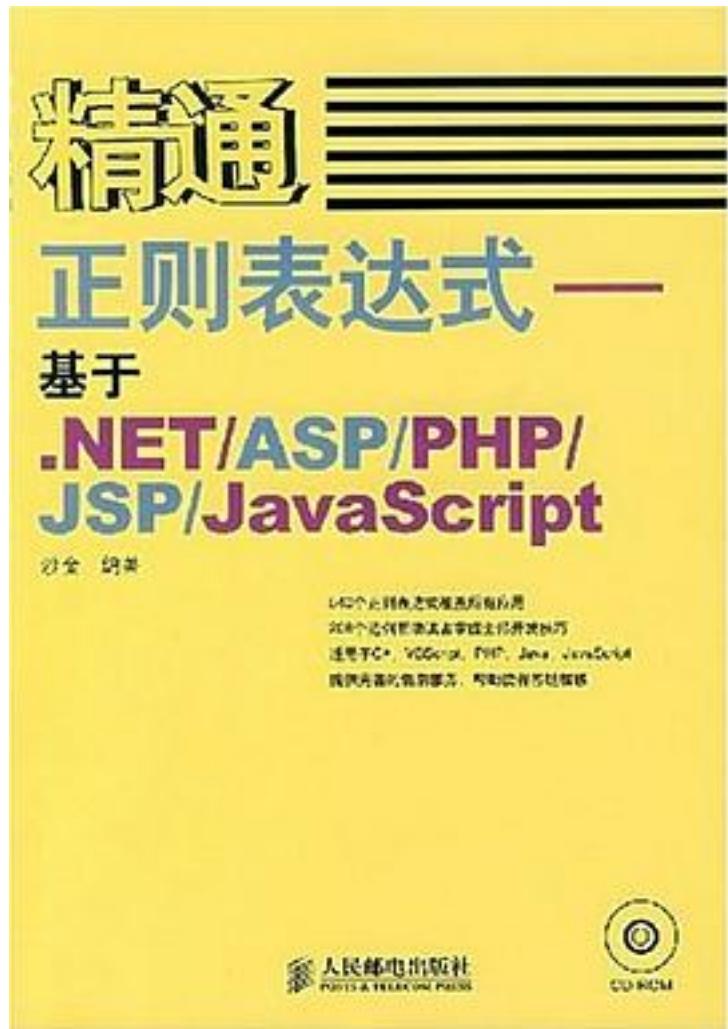


# 精通正则表达式



[精通正则表达式 下载链接1](#)

著者:[美] Jeffrey E · F · Friedl

出版者:电子工业出版社

出版时间:2007

装帧:平装

isbn:9787121046841

随着互联网的迅速发展，几乎所有工具软件和程序语言都支持的正则表达式也变得越来

越强大和易于使用。本书是讲解正则表达式的经典之作。本书主要讲解了正则表达式的特性和流派、匹配原理、优化原则、实用诀窍以及调校措施，并详细介绍了正则表达式在Perl、Java、.NET、PHP中的用法。

本书自第1

版开始着力于教会读者“以正则表达式来思考”，来让读者真正“精通”正则表达式。该版对PHP的相关内容、Java1.5和Java1.6的新特性作了可观的扩充讲解。任何有机会使用正则表达式的读者都会从中获益匪浅。

作者介绍：

目录: 前言

第1章 正则表达式入门

解决实际问题

作为编程语言的正则表达式

以文件名做类比

以语言做类比

正则表达式的思维框架

对于有部分经验的读者

检索文本文件：Egrep

Egrep元字符

行的起始和结束

字符组

用点号匹配任意字符

多选结构

忽略大小写

单词分界符

小结

可选项元素

其他量词：重复出现

括号及反向引用

神奇的转义

基础知识拓展

语言的差异

正则表达式的目标

更多的例子

正则表达式术语汇总

改进现状

总结

一家之言

第2章 入门示例拓展

关于这些例子

Perl简单入门

使用正则表达式匹配文本

向实用的程序前进

成功匹配的副作用

错综复杂的正则表达式

暂停片刻

使用正则表达式修改文本

例子：公函生成程序

举例：修整股票价格

自动的编辑操作

处理邮件的小工具

用环视功能为数值添加逗号

Text-to-HTML转换

回到单词重复问题

第3章正则表达式的特性和流派概览

在正则的世界中漫步

正则表达式的起源

最初印象

正则表达式的注意事项和处理方式

集成式处理

程序式处理和面向对象式处理

查找和替换

其他语言中的查找和替换

注意事项和处理方式：小结

字符串，字符编码和匹配模式

作为正则表达式的字符串

字符编码

Unicode

正则模式和匹配模式

常用的元字符和特性

字符表示法

字符组及相关结构

锚点及其他“零长度断言”

注释和模式量词

分组，捕获，条件判断和控制

高级话题引导

第4章：表达式的匹配原理

发动引擎

两类引擎

新的标准

正则引擎的分类

几句题外话

测试引擎的类型

匹配的基础

关于范例

规则1：优先选择最左端的匹配结果

引擎的构造

规则2：标准量词是匹配优先的

表达式主导与文本主导

NFA引擎：表达式主导

DFA引擎：文本主导

第一想法：比较NFA与DFA

回溯

真实世界中的例子：面包屑

回溯的两个要点

备用状态

回溯与匹配优先

关于匹配优先和回溯的更多内容

匹配优先的问题

多字符“引文”

使用忽略优先量词

匹配优先和忽略优先都期望获得匹配

匹配优先、忽略优先和回溯的要旨

占有优先量词和固化分组

占有优先量词，`?+`、`*+`、`++`和`{m,n}+`

环视的回溯

多选结构也是匹配优先的吗

发掘有序多选结构的价值

NFA、DFA和POSIX

最左最长规则

POSIX和最左最长规则

速度和效率

小结：NFA与DFA的比较

总结

第5章：正则表达式实用技巧

正则表达式的平衡法则

若干简单的例子

匹配连续行（续前）

匹配IP地址

处理文件名

匹配对称的括号

防备不期望的匹配

匹配分隔符之内的文本

了解数据，做出假设

去除文本首尾的空白字符

HTML相关范例

匹配HTMLTag

匹配HTMLLink

检查HTTPURL

验证主机名

在真实世界中提取URL

扩展的例子

保持数据的协调性

解析CSV文件

第6章：打造高效正则表达式

典型示例

稍加修改——先迈最好使的腿

效率vs准确性

继续前进——限制匹配优先的作用范围

实测

全面考查回溯

POSIXNFA需要更多处理

无法匹配时必须进行的工作

看清楚一点

多选结构的代价可能很高

性能测试

理解测量对象

PHP测试

Java测试

VB.NET测试

Ruby测试

Python测试

Tcl测试

常见优化措施

有得必有失

优化各有不同

正则表达式的应用原理

应用之前的优化措施

通过传动装置进行优化  
优化正则表达式本身  
提高表达式速度的诀窍  
常识性优化  
将文字文本独立出来  
将锚点独立出来  
忽略优先还是匹配优先？具体情况具体分析  
拆分正则表达式  
模拟开头字符识别  
使用固化分组和占有优先量词  
主导引擎的匹配  
消除循环  
方法1：依据经验构建正则表达式  
真正的“消除循环”解法  
方法2：自顶向下的视角  
方法3：匹配主机名  
观察  
使用固化分组和占有优先量词  
简单的消除循环的例子  
消除C语言注释匹配的循环  
流畅运转的表达式  
引导匹配的工具  
引导良好的正则表达式速度很快完工  
总结：开动你的大脑  
第7章：Perl  
作为语言组件的正则表达式  
Perl的长处  
Perl的短处  
Perl的正则流派  
正则运算符和正则文字  
正则文字的解析方式  
正则修饰符  
正则表达式相关的Perl教义  
表达式应用场合  
动态作用域及正则匹配效应  
匹配修改的特殊变量  
qr//运算符与regex对象  
构建和使用regex对象  
探究regex对象  
用regex对象提高效率  
Match运算符  
Match的正则运算元  
指定目标运算元  
Match运算符的不同用途  
迭代匹配：ScalarContext，不使用/g  
Match运算符与环境的关系  
Substitution运算符  
运算元replacement  
/e修饰符  
应用场合与返回值  
Split运算符  
Split基础知识  
返回空元素  
Split中的特殊Regex运算元

Split中带捕获型括号的match运算元  
巧用Perl的专有特性  
用动态正则表达式结构匹配嵌套结构  
使用内嵌代码结构  
在内嵌代码结构中使用local函数  
关于内嵌代码和my变量的忠告  
使用内嵌代码匹配嵌套结构  
正则文字重载  
正则文字重载的问题  
模拟命名捕获  
效率  
办法不只一种  
表达式编译、/o修饰符、qr/ . . . /和效率  
理解“原文”副本  
Study函数  
性能测试  
正则表达式调试信息  
结语  
第8章：Java  
Java的正则流派  
Java对\p{}和\P{}的支持  
Unicode行终结符  
使用java.util.regex  
ThePattern.compile()Factory  
Pattern的matcher方法  
Matcher对象  
应用正则表达式  
查询匹配结果  
简单查找-替换  
高级查找-替换  
原地查找-替换  
Matcher的检索范围  
方法链  
构建扫描程序  
Matcher的其他方法  
Pattern的其他方法  
Pattern的split方法，单个参数  
Pattern的split方法，两个参数  
拓展示例  
为ImageTag添加宽度和高度属性  
对于每个Matcher，使用多个Pattern校验HTML  
解析CSV文档  
Java版本差异  
1.4.2和1.5.0之间的差异  
1.5.0和1.6之间的差异  
第9章：.NET  
.NET的正则流派  
对于流派的补充  
使用.NET正则表达式  
正则表达式快速入门  
包概览  
核心对象概览  
核心对象详解  
创建Regex对象

使用Regex对象  
使用Match对象  
使用Group对象  
静态“便捷”函数  
正则表达式缓存  
辅助函数  
.NET高级话题  
正则表达式装配件  
匹配嵌套结构  
Capture对象  
第10章：PHP  
PHP的正则流派  
Preg函数接口  
“Pattern”参数  
Preg函数罗列  
preg\_match  
preg\_match\_all  
preg\_replace  
preg\_replace\_callback  
preg\_split  
preg\_grep  
preg\_quote  
“缺失”的preg函数  
preg\_regex\_to\_pattern  
对未知的Pattern参数进行语法检查  
对未知正则表达式进行语法检查  
递归的正则表达式  
匹配嵌套括号内的文本  
不能回溯到递归调用之内  
匹配一组嵌套的括号  
PHP效率  
模式修饰符S：“研究”  
扩展示例  
用PHP解析CSV  
检查taggeddata的嵌套正确性  
索引  
· · · · · (收起)

[精通正则表达式](#) [下载链接1](#)

标签

正则表达式

编程

计算机

regex

正则

Programming

精通正则表达式第三版

编程基础

## 评论

经典就在于你每次翻阅，几乎都能发现你以前没有留意，但是现在突然意识到的那些问题的答案。

---

: TP301.2/2212

---

省钱了 - - - 感谢图书馆.....

---

翻译很不认真，很多正则表达式原封不动的复制过来都能整错。  
本书好是好，但是新手第一遍很难完全理解。

---

不会正则的程序员，都不敢大声说话

---

用正则表达式很容易，要达到精通的境界就要花些功夫了~  
弄懂正则表达式的工作原理才是真正理解的关键，本书就详细地介绍了走向精通的知识点~ 翻译还不错，几处纠结的地方看下原版也就明白了，印刷有若干错误啦~

作者从事多年正则方面的工作，从玩Egrep和perl中的正则到现在各种语言中的正则，对正则的历史变更了如指掌，对此技术是绝对的权威，我强烈建议学习正则的朋友阅读此书！

作者用心良苦，为了引导读者去思考，特意将提问的答案放在下一页！作者以讲故事的形式来写作此书，以大白话来讲述技术，文字功底很好，翻译的质量也是上层！

此书浅显易懂，读起来兴趣盎然！

没认真看

余晨的心血译作。

（文本）字符串的数据类型的操作的方法--形式代数-不懂英语（单词，句子，大意），但是仅仅理解字节和行。正则表达式的语法，实际上是一种轻量级、简洁、适用于特定领域的编程语言

看了前六章，剩下的没有读完

#纸质书# 还蛮不错的吧~ 正则原来这么丰富，但是真的用得到这么么？

manual

.....我自卑啊.....

很基础

讲到还不错，如果总写的话，语法很熟，最近不写了，忘的也差不多了

读了前面部分，对正则表达式的认识有所提升。

一本特别好的书，特别有意思。但不一定需要我认真读下去==方向不同罢了，但是还是推荐给各位发烧友们……大神的世界我还是不懂得为妙

一本由满足够用向专家过渡的专业图书，如果你只是满足一般的编程需要，没必要如此精通。但一个精通正则表达式的程序员力量更加强大。

太艰深 看不下去。。

[精通正则表达式 下载链接1](#)

## 书评

如果你只想“用”正则表达式，那么就别读这本书了，市面上任何一本编程书籍中讲解正则表达式的一章就足够了，网上的资料也很齐全。本书面向的读者是 1) 会用正则表达式 2) 愿意从一个代码工人向专家进化的 3) 对技术有狂热的追求的。本书着重讲解关于正则表达式匹配原理...

Mastering Regular Expressions 3rd Edition.pdf  
CSDN版本，配合一个勘误表来看。一共有10章，建议看前五章，各个章节的收获如下：  
看完1,2章：解决80%的实际问题  
看完第3章：90%  
看完第4章：97%  
看完第5章：99.9%  
第6--10章不要看，原因很显而易见，因为99.9%的能够用...

详见<http://blog.donews.com/maverick/archive/2007/09/14/1209937.aspx>

先说说看完这本书之后，我对正则表达式的重新认识：正则表达式就是查字典。当然，并不是说真的就是查字典，只是过程很类似而已。举个例子：当你要查询以R开头的单词的时候，这个问题就可能用不到正则表达式；但是当你要查询第三个字母是R的单词的时候，用正则...

这个本买来很长时间了 感觉太厚 就一直没有看，这点时间突然对正则表达式感兴趣了就看了前几张的内容很值得看 尤其是中间几章，我是彻底了解了正则的工作原理以前工作中出现的一些情况原来想不明白，现在也就觉得应该是这样子了建议大家静下心来看这本书

若论入门，《正则表达式必知必会》之类的书比这本书讲得简明易懂。若单说Perl的正则表达式，《Progammimg Perl》比本书讲得权威。这本书涵括的语言包括Perl、Java、.NET、PHP，但是我相信很少有人同时用这几种语言编程的。我以为本书比较有价值的部分是第6章：打造高效的正则表...

虽说看过manual，几个例子，每个人都可以开始写regex了但是磨刀不误砍柴工，还是把原理搞清楚，以后可以省去你调试的大量时间，最重要的是你可以在那些只知道basic的菜鸟面前显摆NFA,DFA,POSIX NFA，驱动，回溯的道理里面有一个词汇翻译的特别不好greedy，有限匹配，贪婪...

正则式，以前看到这本书时就感觉正则式必然是非常高深的技术，但是学习了编译原理-词法分析以及有限自动机后，便觉得正则式不再是高深而抽象。建议读者本书前花一点时间学习一下广为流传的速成教程：正则表达式30分钟入门教程，然后再读一读这本书。（尽管这本书的第一章号称...

不管出于何种原因，不管你的目标是只是达到能用，还是希望成为正则大师(如书名mastering regular expression)，读这本书，且只这一本书，足矣。该书拥有所有成为技术经典的必备优点，结构编排合理，行文用词准确，技术点涵盖的全面而且深入。这本书大致分三个部分，第一部分...

这本书讲了不少关于正则表达式引擎的东西，并且花费了很大力气讲解基于回溯的NFA引擎。好像回溯是实现正则引擎的唯一算法。

事实上，有很多更高效的算法，我自己就实现过一个正则引擎，专门针对正则表达式集合的匹配，也就是说，给定很多个正则表达式（比如100万个），对输入...

书是好书，如之前一人的评论，性价比很高。个人认为，如果你的日常工作经常涉及文本处理，那读本书应该能提升你的工作效率。

再谈翻译，虽然很多人说，本书翻译的质量不错，但实际上，翻译有很多错误，而且一些错误直接影响到文义，如要精读，还是应该辅以英文原文。

首先术语翻译不佳，扣掉一星。例如atomic

grouping翻译为“固化分组”，就是“原子分组”吗。其他“匹配优先”就是“贪婪“匹配吗，还有一些别的，看的很不顺畅。

其次，原文介绍了正则表达式的应用和使用技巧以及回溯等原理性介绍，但是没有介绍code是如何实现的，因此有些地...

正则表达式非常强大，在各种语言都有相关的包或库。看完了前五章已经对我大有帮助！翻译也不错，读起来顺畅，内容包括了从入门到精通，后面几章主要是分语言来讲解正则的应用，有Perl、Java、.NET和PHP，可以选自己掌握的语言来看。本人认为是程序员必读的一本工具书。

看正则的缘由：Vim用的有一短时间了，

时常用里面的替换功能，但有时发现简单替换仍旧力不从心，觉得应该看些正则的东西

。网上查了些正则的例子，但看过例子之后，只是觉得当前的东西可以用了，却始终抓不到边际本质。暑假决定给自己放一周，平时的MRI的东西和Matlab，Pyth...

这是我最近读过些的最好的技术书籍了。原来正做表达式的用法可以这么神奇，虽然原来也一直是必备工具，但是从来没想到过怎么在整篇文章中直接用一个很简单的表达式找到连续重复的单词！想知道吗？去看看原著吧，就在第一章里面。

书比我想象中的厚，当初选择看这本书完全是被电子书的大小给骗了，网上流传的epub格式的电子书虽然1M不到，本还打算当做厕所读物看的，结果没想到这本书看了这么

久，久到我刷完了《good luck charlie》四季，这本书还在坚挺着摆在我的书桌上。另外在这里提一句，千万别把网上流...

---

---

第一类：GEN-ADML-ST-IDENTIFIANT.XXXX 第二类：S2131000\_XXX\_REQ\_XXX  
注：XXX表示数字 第一类是用 ‘-’ 第二类是用 ‘\_’

---

佩服作者，真的是很厉害，佩服他对各种语言的了解和掌握。  
正则表达式对于很多我遇到的新手或者高手都是他们的一个瓶颈，有很多人不了解或者不是很了解正则表达式，其实如果不是之前的工作逼迫我也不会接触它，但是真的上手以后发现，正则真是一个好东西，简洁优美。但是查询...

---

[精通正则表达式](#) [下载链接1](#)