第一图书网, tushu007.com

<<正则表达式经典实例>>

图书基本信息

书名：<<正则表达式经典实例>>

13位ISBN编号：9787115228321

10位ISBN编号：7115228329

出版时间：2010-6

出版时间：人民邮电出版社

作者：[美]高瓦特斯 (Goyvaerts, Jan),[美]利维森 (Levithan, Steven)

页数：457

译者：郭耀

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：http://www.tushu007.com

第一图书网, tushu007.com

前言

正则表达式在过去十多年间越来越普及。
如今所有常用的编程语言都会包含一个强大的正则表达式函数库，或者甚至是在语言本身就内嵌了对于正则表达式的支持。
许多开发人员都会利用这些正则表达式的功能，在应用程序中为用户提供使用正则表达式对其数据进行查找或者过滤的能力。
正则表达式真正是无处不在。
随着正则表达式的广泛采用，出现了许多相关的著作。
大多数这类书籍都很好地讲解了正则表达式的语法，并且还会提供一些例子以及参考文献。
然而，我们还没有看到有任何一本书能够面向现实世界中使用的计算机，以及在各种：Internet应用中遇到的实际问题，为读者提供基于正则表达式的解决方案。
因此，本书作者Steve和Jan决定写一本书来填补这种空白。
我们特别期望能够展现给读者的是：如何使用正则表达式来解决那些对正则表达式经验较少的人们认为无法解决的问题，以及软件纯粹主义者认为不能用正则表达式来解决的问题。
因为如今正则表达式无处不在，所以它们通常可以作为工具，直接被最终用户使用，而不需要程序员团队的参与。
即使是对于程序员来说，常常也可以在信息检索和更新的任务中采用一些正则表达式来节省大量时间，因为这些功能如果使用过程式代码来实现，可能就会需要几个小时甚至几天的时间，也可能会由于需要采用第三方的函数库，而不得不经过事先审查和经理层的审批。
不同版本带来的问题与IT业界流行的东西一样，正则表达式也拥有许多种不同的实现，以及不同程度的兼容性。
这就出现了许多不同的正则表达式流派（flavor），它们在处理一个特定正则表达式的时候并不总是拥有完全一样的表现，有时候甚至会无法正常使用。
在许多书中的确也提到了目前存在的不同流派，并且指出了其中的一些区别。
但是，如果某种流派缺少特定功能的时候，它们通常会选择在这里或那里略掉一些流派，而不是为之提供可替代的解决方案或者是应急方案。
而当你不得不在不同的应用程序或者不同程序语言中使用不同的正则表达式流派的时候，就会感到很受挫折。

第一图书网, tushu007.com

<<正则表达式经典实例>>

内容概要

本书讲解了基于8种常用的编程语言使用正则表达式的经典实例。
书中提供了上百种可以在实战中使用的实例，以帮助读者使用正则表达式来处理数据和文本。
对于如何使用正则表达式来解决性能不佳、误报、漏报等常见的错误以及完成一些常见的任务，本书给出了涉及基于C#、Java、JavaScript、Perl、PHP、Python、Ruby和VB.NET等编程语言的解决方案。

本书的读者对象是对正则表达式感兴趣的软件开发人员和系统管理员。
本书旨在教会读者很多新的技巧以及如何避免语言特定的陷阱，读者可以通过本书提供的实例解决方案库来解决实践中的复杂问题。

第一图书网, tushu007.com

作者简介

Jan Goyvaerts领导着Just Great Software公司，他在这个公司设计和开发了一些最流行的正则表达式软件。
他的产品中包括RegexBuddy，世界上唯一可以模拟l5种正则表达式流派特性的正则表达式编辑器，以及PowerGREP，Microsoft Windows平台上功能最强大的9rep工具。

第一图书网, tushu007.com

<<正则表达式经典实例>>

书籍目录

第1章 正则表达式简介 1.1 正则表达式的定义 1.2 使用正则表达式的工具第2章 正则表达式的基本技巧 2.1 匹配字面文本 2.2 匹配不可打印字符 2.3 匹配多个字符之一 2.4 匹配任意字符 2.5 匹配文本行起始和/或文本行结尾 2.6 匹配整个单词 2.7 Unicode代码点、属性、区块和脚本 2.8 匹配多个选择分支之一 2.9 分组和捕获匹配中的子串 2.10 再次匹配先前匹配的文本 2.11 捕获和命名匹配子串 2.12 把正则表达式的一部分重复多次 2.13 选择最小和最大重复次数 2.14 消除不必要的回溯 2.15 避免重复逃逸 2.16 检查一个匹配，但不添加到整体匹配中 2.17 根据条件匹配两者之一 2.18 向正则表达式中添加注释 2.19 在替代文本中添加字面文本 2.20 在替代文本中添加正则匹配 2.21 把部分的正则匹配添加到替代文本中 2.22 把匹配上下文插入到替代文本中第3章 使用正则表达式编程 3.1 在源代码中使用字面正则表达式 3.2 导入正则表达式函数库 3.3 创建正则表达式对象 3.4 设置正则表达式选项 3.5 检查是否可以在目标字符串中找到匹配 3.6 检查正则表达式能否整个匹配目标字符串 3.7 获取匹配文本 3.8 决定匹配的位置和长度 3.9 获取匹配文本的一部分 3.10 获取所有匹配的列表 3.11 遍历所有匹配 3.12 在过程代码中对匹配结果进行验证 3.13 在另一个匹配中查找匹配 3.14 替换所有匹配 3.15 使用匹配的子串来替换匹配 3.16 使用代码中生成的替代文本来替换匹配 3.17 替换另一个正则式匹配中的所有匹配 3.18 替换另一个正则式匹配之间的所有匹配 3.19 拆分字符串 3.20 拆分字符串，保留正则匹配 3.21 逐行查找第4章 合法性验证和格式化 4.1 E-mail地址的合法性验证 4.2 北美电话号码的合法性验证和格式化 4.3 国际电话号码的合法性验证 4.4 传统日期格式的合法性验证 4.5 对传统日期格式进行精确的合法性验证 4.6 传统时间格式的合法性验证 4.7 检查ISO 8601格式的日期和时间 4.8 限制输入只能为字母数字字符 4.9 限制文本长度 4.10 限制文本中的行数 4.11 肯定响应的检查 4.12 社会安全号码的合法性验证 4.13 ISBN的合法性验证 4.14 ZIP代码的合法性验证 4.15 加拿大邮政编码的合法性验证 4.16 英国邮政编码的合法性验证 4.17 查找使用邮局信箱的地址 4.18 转换姓名格式 4.19 信用卡号码的合法性验证 4.20 欧盟增值税代码第5章 单词、文本行和特殊字符 5.1 查找一个特定单词 5.2 查找多个单词之一 5.3 查找相似单词 5.4 查找除某个单词之外的任意单词 5.5 查找后面不跟着某个特定单词的任意单词 5.6 查找不跟在某个特定单词之后的任意单词 5.7 查找临近单词 5.8 查找重复单词 5.9 删除重复的文本行 5.10 匹配包含某个单词的整行内容 5.11 匹配不包含某个单词的整行 5.12 删除前导和拖尾的空格 5.13 把重复的空白替换为单个空格 5.14 对正则表达式元字符进行转义第6章 数字 6.1 整数 6.2 十六进制数字 6.3 二进制数 6.4 删除前导 6.5 位于某个特定范围之内的整数 6.6 在某个特定范围之内的十六进制数 6.7 浮点数 6.8 含有千位分隔符的数 6.9 罗马数字第7章 URL、路径和Internet地址 7.1 URL合法性验证 7.2 在全文中查找URL 7.3 在全文中查找加引号的URL 7.4 在全文中寻找加括号的URL 7.5 把URL转变为链接 7.6 URN合法性验证 7.7 通用URL的合法性验证 7.8 从URL中提取通信协议方案 7.9 从URL中抽取用户名 7.10 从URL中抽取主机名 7.11 从URL中抽取端口号 7.12 从URL中抽取路径 7.13 从URL中抽取查询 7.14 从URL中抽取片段 7.15 域名合法性验证 7.16 匹配IPv4地址 7.17 匹配IPv6地址 7.18 Windows路径的合法性验证 7.19 分解Windows路径 7.20 从Windows路径中抽取盘符 7.21 从UNC路径中抽取服务器和共享名 7.22 从Windows路径中抽取文件夹 7.23 从Windows路径中抽取文件名 7.24 从Windows路径中抽取文件扩展名 7.25 去除文件名中的非法字符 第8章 标记语言和数据交换 8.1 查找XML风格的标签 8.2 把标签b替换为strong 8.3 删掉除em和strong之外的所有XML风格标签 8.4 匹配XML名称 8.5 添加p和br标签将纯文本转换为HTML 8.6 在XML风格的标签中查找某个特定属性 8.7 向不包含cellspacing属性的table标签中添加该属性 8.8 删除XML风格的注释 8.9 在XML风格的注释中查找单词 8.10 替换在CSV文件中使用的分隔符 8.11 抽取某个特定列中的CSV域 8.12 匹配INI段头 8.13 匹配INI段块 8.14 匹配INI名称-值对

第一图书网, tushu007.com

章节摘录

插图：在这最后一章中，我们会着重讲解在处理常见的标记语言时可能会遇到的常见任务，这些标记语言包括：HTMI、XHTML、XML、CSV和INI。
虽然我们会假设读者对于这些技术至少已经有了基本的了解，但是在本章开始时，我们还是会对每种语言进行简单的介绍，以确保在进一步讲解之前都站在同一起跑线上。
这里的描述主要关注的是在搜索每种格式的数据结构时所需要知道的基本语法规则。
当我们在后面遇到相关问题的时候，届时会再引人更多的细节。
要想准确处理和操作其中的一些格式虽然从表面上看并不总是很难，然而有时候这样的任务会极其复杂，至少使用正则表达式来处理时是如此。
如果需要完成本章中讲解的许多任务，特别是当在准确性非常关键的时候（例如，如果你的处理过程可能会涉及系统安全），那么通常最好的选择应当是使用专门的分析器和API，而不是采用正则表达式。
然而，这些实例依然介绍了许多有用的技巧，并且可以用于许多需要快速处理的任务。
因此，我们先来看一下要处理的到底是什么东西。
在本章中会遇到的许多困难都涉及如何处理与下列规则不一致的情形，这其中可能是可以预期的情形，也可能是无法预期的情形。

第一图书网, tushu007.com

媒体关注与评论

“这是一本严谨的著作，内容非常丰富。
仅仅阅读前几章我就学到了大量的新技巧。
”　　——Nikolaj Lindberg 计算语言学家sTTs语音技术服务公司“本书为紧迫的问题提供了很好的解决方案。
对于实例中所包含的细节我感到非常震惊。
”　　——Zak Greant 开放技术倡导者和策略专家Jan Goyvaerts领导专家

第一图书网, tushu007.com

<<正则表达式经典实例>>

编辑推荐

《正则表达式经典实例》：每个程序员都会遇到需要使用正则表达式的情况，但是要用好正则表达式却并不容易。
《正则表达式经典实例》提供了100多个实例，以帮助读者使用正则表达式处理数据和文本。
即使有经验的用户也经常会遇到性能不佳、误报、漏报等让人挠头的错误，《正则表达式经典实例》对于如何使用正则表达式来解决一些常见的问题给出了按部就班的解决方案，其中包括C＃、Java、JavaScript、Perl、PHP、Python、Ruby和VB.NET的实例。
《正则表达式经典实例》主要包括以下内容：通过一个精练的教程理解正则表达式的基本原理和技巧；在不同的编程语言和脚本语言中有效使用正则表达式；学习如何对输入进行合法性检查和格式化；处理单词、文本行、特殊字符和数值；学习如何在uRL、路径、标记语言和数据交换中使用正则表达式；学习更高深的正则表达式特性中的微妙之处；理解在不同语言中正则表达式的API、语法和行为之间的区别；创建更好的正则表达式来满足个性化的需求。
使用8种编程语言的详细解决方案,包括一个正则表达式简明教程

第一图书网, tushu007.com

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:http://www.tushu007.com