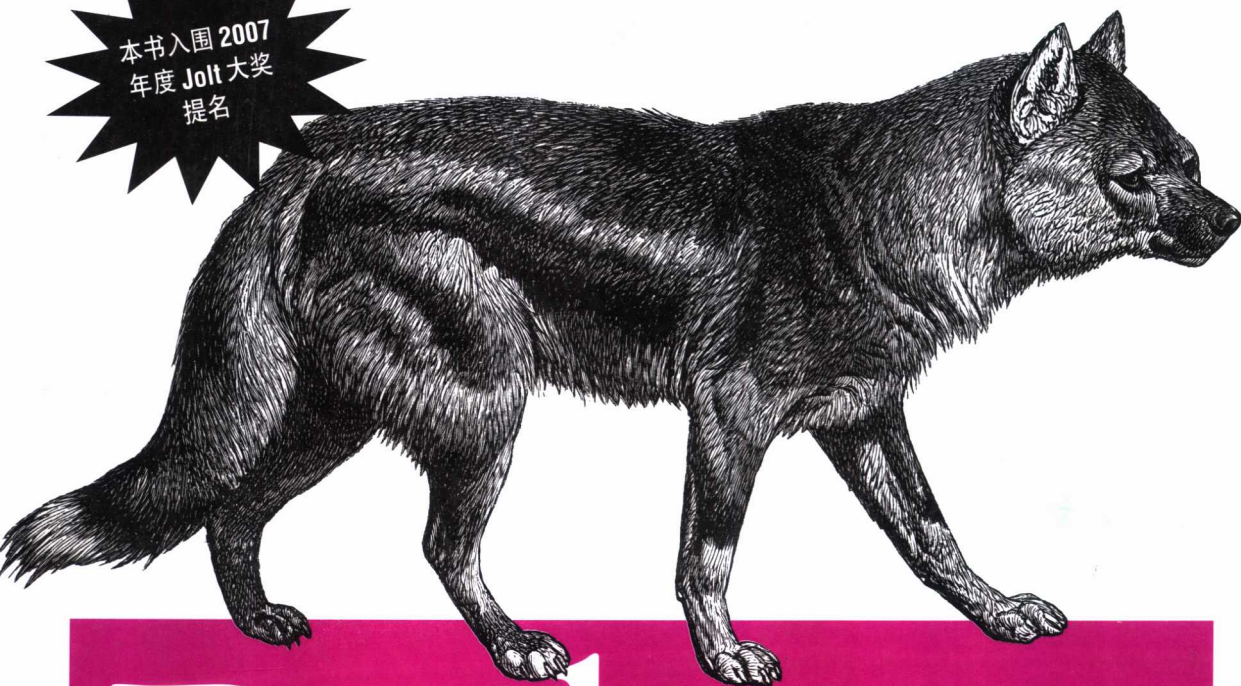


Ruby Cookbook

本书入围 2007
年度 Jolt 大奖
提名



Ruby Cookbook™ 中文版

Lucas Carlson & Leonard Richardson 著
郑路长 译

O'REILLY®



清华大学出版社

Copyright ©2006 by O'Reilly Media, Inc.

Authorized Simplified Chinese translation edition, by O'Reilly Media, Inc., is published by Tsinghua University Press, 2007. Authorized translation of the original English edition, 2006 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

本书之英文原版由 O'Reilly Media, Inc. 于 2006 年出版。

本中文简体翻译版由 O'Reilly Media, Inc. 授权清华大学出版社于 2007 年出版。此翻译版的出版和销售得到出版权和销售权的所有者——O'Reilly Media, Inc. 的许可。

版权所有，未经书面许可，本书的任何部分和全部不得以任何形式复制。

北京市版权局著作权合同登记

图字：01-2007-1165 号

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13501256678 13801310933

图书在版编目 (CIP) 数据

Ruby Cookbook™ 中文版 / (美) 卡尔松 (Carlson, L.), (美) 理查得松 (Richardson, L.) 著; 郑路长译. —北京: 清华大学出版社, 2007.4

书名原文: Ruby Cookbook

ISBN 978-7-302-14770-1

I. R… II. ①卡… ②理… ③郑… III. 软件开发 IV. TP311.52

中国版本图书馆 CIP 数据核字 (2007) 第 026896 号

责任编辑: 常晓波

封面设计: Karen Montgomery, 张 健

责任校对: 张 剑

责任印制: 王秀菊

出版发行: 清华大学出版社 地 址: 北京清华大学学研大厦 A 座

<http://www.tup.com.cn> 邮 编: 100084

c-service@tup.tsinghua.edu.cn

社 总 机: 010-62770175 邮购热线: 010-62786544

投稿咨询: 010-62772015 客户服务: 010-62776969

印 刷 者: 清华大学印刷厂

装 订 者: 三河市金元印装有限公司

经 销: 全国新华书店

开 本: 178 毫米 × 233 毫米 57.75 印张 字数: 1259 千字

版 次: 2007 年 4 月第 1 版 印次: 2007 年 4 月第 1 次印刷

印 数: 1 ~ 3000 册

定 价: 108.00 元(册)

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题, 请与清华大学出版社出版部联系调换。联系电话: (010)62770177 转 3103 产品编号: 023981-01

目录

前言	1
第 1 章 字符串	13
1.1 根据部分构建字符串	16
1.2 将变量代入字符串	18
1.3 将变量代入现有的字符串	20
1.4 按字或字符逆转字符串	22
1.5 表示不可打印字符	23
1.6 字符与值的相互转换	26
1.7 字符串与符号之间的相互转换	27
1.8 每次处理字符串中的一个字符	28
1.9 每次处理字符串中的一个字	30
1.10 修改字符串的大小写	32
1.11 管理空白	34
1.12 测试对象是否类似于字符串	35
1.13 获得想要的字符串部分	36
1.14 处理国际编码	38
1.15 带自动换行的文本	39

1.16	生成字符串的后继	41
1.17	使用正则表达式匹配字符串	44
1.18	一遍替换多种模式	46
1.19	验证电子邮箱地址	47
1.20	使用贝叶斯分析器对文本进行分类	50

第 2 章 数字 53

2.1	由字符串分析数字	54
2.2	浮点数的比较	57
2.3	表示数字至任意精度	59
2.4	有理数的表示	62
2.5	随机数的生成	64
2.6	数字基之间的转换	66
2.7	获取对数	67
2.8	确定平均数、中值和模	69
2.9	度数与弧度之间的转换	72
2.10	矩阵乘法	74
2.11	求解线性方程组	79
2.12	复数的使用	81
2.13	模拟 Fixnum 的子类	84
2.14	使用罗马数字进行数学运算	88
2.15	生成数字序列	93
2.16	生成素数	96
2.17	检查信用卡校验和	99

第 3 章 日期与时间 102

3.1	查找当前日期	105
3.2	精确或模糊地分析日期	108
3.3	打印日期	111
3.4	日期上的迭代	117

3.5	计算日期	119
3.6	从任意日期开始计算天数	121
3.7	时区转换	123
3.8	检查夏令时是否起效	126
3.9	Time 与 DateTime 对象之间的转换	127
3.10	查找周日期	130
3.11	处理商用日期	132
3.12	周期性运行代码块	133
3.13	等待固定长度的时间	135
3.14	为长期运行的操作添加超时	138
 第 4 章 数组		140
4.1	在数组上迭代	142
4.2	不使用临时变量重排值	146
4.3	去除数组中的重复元素	148
4.4	逆转数组	149
4.5	数组排序	150
4.6	排序字符串时忽略大小写	152
4.7	确保已排序数组保持有序	153
4.8	数组项求和	158
4.9	按出现率排序数组	159
4.10	打乱数组	161
4.11	获取数组的 N 个最小项	163
4.12	使用 Injection 构建散列	165
4.13	提取部分数组	167
4.14	在数组上处理集合运算	170
4.15	集合划分或分类	173
 第 5 章 散列		177
5.1	使用符号作为散列键	179

5.2	创建带默认值的散列	181
5.3	向散列添加元素	182
5.4	从散列中去除元素	185
5.5	使用数组或其他可修改对象作为散列键	186
5.6	为相同散列键保持重复值	188
5.7	在散列上迭代	190
5.8	按插入顺序在散列上迭代	192
5.9	散列打印	193
5.10	反转散列	196
5.11	随机选择加权列表	197
5.12	构建柱状图	200
5.13	重新映射散列的键与值	202
5.14	提取部分散列	203
5.15	使用正则表达式搜索散列	204

第 6 章 文件与目录 206

6.1	检查文件是否存在	209
6.2	检查对文件的访问	210
6.3	更改文件权限	212
6.4	查看上次使用文件的时间	216
6.5	目录清单	217
6.6	读取文件内容	220
6.7	写文件	224
6.8	写临时文件	225
6.9	从文件中挑选随机行	227
6.10	比较两个文件	228
6.11	在“只读一次”输入流上执行随机访问	232
6.12	遍历目录树	234
6.13	文件加锁	237
6.14	备份至带版本号的文件名	240

6.15	伪装字符串为文件	242
6.16	重定向标准输出或输出	245
6.17	处理二进制文件	247
6.18	删除文件	251
6.19	截短文件	252
6.20	查找所需文件	253
6.21	查找并修改当前工作目录	255
 第 7 章 代码块与迭代		257
7.1	代码块的创建与调用	260
7.2	编写接收代码块的方法	262
7.3	将代码块参数与变量绑定	264
7.4	作为闭包的代码块：在代码块内使用外部变量	267
7.5	在数据结构上编写迭代器	268
7.6	更改对象的迭代方式	271
7.7	编写分类或收集的代码块方法	273
7.8	停止迭代	275
7.9	并行地在多个迭代变量上循环	277
7.10	隐藏块方法中的设置与清除	281
7.11	使用回调的松耦合系统	284
 第 8 章 对象与类		288
8.1	管理实例数据	290
8.2	管理类数据	293
8.3	检查类或模块的成员	296
8.4	编写一个继承类	298
8.5	方法重载	300
8.6	验证并修改属性值	302
8.7	定义虚属性	304
8.8	授权对另一对象的方法调用	305

8.9	对象到不同类型的转换与强制转换	308
8.10	从任意对象获取人类可读的打印输出	312
8.11	接收或传递参数的变量数目	315
8.12	模拟关键字参数	317
8.13	调用超类的方法	319
8.14	创建抽象方法	321
8.15	冻结对象以防修改	324
8.16	生成对象的副本	326
8.17	声明常量	329
8.18	实现类方法和 singleton 方法	331
8.19	通过私有化方法控制访问	333

第 9 章 模块与命名空间 337

9.1	使用混入模拟多重继承	337
9.2	使用模块扩展特定对象	341
9.3	混用类方法	343
9.4	实现 Enumerable: 编写一个方法, 获得 22 种免费方法	345
9.5	使用命名空间避免名字冲突	347
9.6	按需自动加载库	349
9.7	包括命名空间	351
9.8	初始化模块定义的实例变量	352
9.9	自动初始化混合插入的模块	353

第 10 章 反射与元编程 357

10.1	查找对象的类和超类	358
10.2	列出对象的方法	359
10.3	列出对象独有的方法	362
10.4	获得方法的引用	363
10.5	修正别人类中的错误	365
10.6	侦听类的变化	367

10.7	检查对象是否具有必需的属性	369
10.8	响应对未定义方法的调用	371
10.9	自动初始化实例变量	375
10.10	使用元编程避免刻板代码	377
10.11	带字符串计算的元编程	379
10.12	计算早先上下文中的代码	382
10.13	取消定义方法	383
10.14	为方法起别名	386
10.15	面向方面的编程	389
10.16	强制实施软件契约	391

第 11 章 XML 和 HTML 396

11.1	检查 XML 的良构性	397
11.2	从文档的树结构中提取数据	399
11.3	解析文档时提取数据	401
11.4	使用 XPath 导航文档	403
11.5	解析不合法标记	405
11.6	将一个 XML 文档转换为一个散列	408
11.7	验证 XML 文档	411
11.8	取代 XML 实体	414
11.9	创建并修改 XML 文档	417
11.10	压缩 XML 文档中的空白	420
11.11	猜解文档的编码	422
11.12	从一种编码转换为另一种编码	423
11.13	从 HTML 文档中提取所有 URL	424
11.14	将纯文本转换为 HTML	428
11.15	将 HTML 文档从 Web 转换为文本	429
11.16	一个简单的提要聚合器	432

第 12 章 图形与其他文件格式..... 437

12.1	缩略图形	437
12.2	向图形中添加文本	440
12.3	将一个图形格式转换为另一种	443
12.4	用图表示数据	446
12.5	使用 Sparkline 添加图形化上下文	449
12.6	强加密数据	453
12.7	解析逗号分隔的数据	455
12.8	解析非完全逗号分隔的数据	457
12.9	生成并分析 Excel 电子数据表	459
12.10	使用 Gzip 和 Tar 压缩并存档文件	462
12.11	读写 ZIP 文件	465
12.12	读写配置文件	467
12.13	生成 PDF 文件	468
12.14	将数据表示为 MIDI 音乐	473

第 13 章 数据库和持久性..... 477

13.1	用 YAML 串行化数据	480
13.2	用 Marshal 串行化数据	484
13.3	用 Madeleine 保持对象	485
13.4	用 SimpleSearch 索引结构化文本	488
13.5	用 Ferret 索引结构化文本	490
13.6	使用 Berkeley DB 数据库	493
13.7	在 Unix 上控制 MySQL	495
13.8	找到查询返回的行数	496
13.9	与 MySQL 数据库直接对话	499
13.10	和 PostgreSQL 数据库直接对话	501
13.11	用 ActiveRecord 使用对象相关映射	504
13.12	使用对象相关映射 Og	508
13.13	以编程方式构建查询	512

13.14 用 ActiveRecord 确认数据	516
13.15 阻止 SQL 注入攻击	519
13.16 在 ActiveRecord 里使用交易	522
13.17 添加挂钩程序到表事件中	524
13.18 用数据库 Minxin 添加标签	526

第 14 章 Internet 服务 531

14.1 抓取 Web 页面的内容	532
14.2 发送 HTTPS Web 请求	534
14.3 自定义 HTTP 的请求头文件	536
14.4 执行 DNS 查询	539
14.5 发送邮件	541
14.6 用 IMAP 阅读邮件	545
14.7 用 POP3 阅读邮件	549
14.8 作为 FTP 客户端	552
14.9 作为 Telnet 客户端	554
14.10 作为 SSH 客户端	558
14.11 复制文件到其他机器	560
14.12 作为 BitTorrent 客户端	561
14.13 ping 机器	564
14.14 编写 Internet 服务器	565
14.15 分析 URL	567
14.16 编写 CGI 脚本	570
14.17 设置 cookie 和其他 HTTP 响应头文件	573
14.18 用 CGI 处理文件上传	575
14.19 用 WEBrick 运行 servlet	578
14.20 真实世界的 HTTP 客户端	584

第 15 章 Web 开发：Ruby on Rails 588

15.1 编写简单的 Rails 应用程序显示系统状态	590
-----------------------------------	-----

15.2 从控制器传递数据到视图	593
15.3 创建页眉和页脚的布局	596
15.4 重新定位不同的位置	599
15.5 用 render 显示模板	601
15.6 集成数据库到 Rails 应用程序中	604
15.7 理解复数规则	607
15.8 创建登录系统	609
15.9 保存散列化的用户密码到数据库中	613
15.10 转义显示用的 HTML 和 JavaScript	615
15.11 设置并找回会话信息	616
15.12 设置并找回 Cookie	619
15.13 提取代码到辅助函数中	621
15.14 重构视图为视图的部分片断	623
15.15 用 script.aculo.us 添加 DHTML 效果	627
15.16 生成操作模型对象的表格	629
15.17 创建 Ajax 表格	633
15.18 在 Web 站点上发布 Web 服务	636
15.19 用 Rails 发送邮件	639
15.20 自动发送错误信息到邮箱	641
15.21 文档化 Web 站点	643
15.22 Web 站点的单元测试	645
15.23 在 Web 应用程序中使用断点	648
第 16 章 Web 服务及分布式编程	652
16.1 搜索 Amazon 上的书	653
16.2 找到 Flickr 上的照片	656
16.3 编写 XML-RPC 客户端	659
16.4 编写 SOAP 客户端	662
16.5 编写 SOAP 服务器	663
16.6 用 Google 的 SOAP 服务搜索 Web	665

16.7	使用 WSDL 文件更简单地构建 SOAP 调用	667
16.8	用信用卡支付	669
16.9	通过 UPS 或 FedEx 找到装运包的费用	671
16.10	在任意数目的计算机间共享散列	672
16.11	实现分布式查询	676
16.12	创建共享的“白板”	678
16.13	通过访问控制列表保障 DRb 服务的安全	682
16.14	通过 Rinda 自动发现 DRb 服务	683
16.15	代理无法分布的对象	685
16.16	用 MemCached 在分布式 RAM 上保存数据	688
16.17	用 MemCached 高速缓存重要结果	690
16.18	远程控制的 Jukebox	694
 第 17 章 测试、调试、优化以及文档化		700
17.1	只在调试模式下运行代码	701
17.2	发出异常	703
17.3	处理异常	705
17.4	在异常后重新运行	707
17.5	添加日志到应用程序中	709
17.6	创建并理解 Traceback	711
17.7	编写单元测试	714
17.8	运行单元测试	717
17.9	测试使用外部资源的代码	720
17.10	使用断点审查并改变应用程序的状态	724
17.11	文档化应用程序	727
17.12	记录应用程序	731
17.13	Benchmark 竞争性解决方案	734
17.14	一次运行多个分析工具	736
17.15	谁调用了该方法? 调用图形分析器	738

第 18 章 打包和发布软件	741
18.1 通过查询 gem 知识库寻找库	742
18.2 安装并使用 gem	745
18.3 要求 gem 的某个特定版本	748
18.4 卸载 gem	751
18.5 为已安装的 gem 读入文档	752
18.6 打包代码为 gem	754
18.7 发布 gem	757
18.8 用 setup.rb 安装并创建独立 Ruby 程序包	760
 第 19 章 用 Rake 自动执行任务	 763
19.1 自动运行单元测试	765
19.3 清除生成的文件	770
19.4 自动构建 gem	771
19.5 收集代码的统计信息	773
19.6 发布文档	776
19.7 并行运行多个任务	778
19.8 通用的项目 Rakefile	779
 第 20 章 多任务和多线程	 787
20.1 在 Unix 上运行守护进程	788
20.2 创建 Windows 服务	791
20.3 用线程一次做两件事情	795
20.4 同步访问一个对象	797
20.5 中止线程	800
20.6 在很多对象上同时运行代码块	802
20.7 用线程池限制多线程	805
20.8 用 popen 驱动外部进程	808
20.9 通过 Unix shell 命令抓取输出和错误流	810
20.10 控制其他机器上的进程	811

20.11 避免死锁	812
第 21 章 用户界面	816
21.1 一次得到输入中的一行	817
21.2 一次得到输入的一个字符	819
21.3 分析命令行参数	822
21.4 测试程序是否交互运行	825
21.5 设置和卸载 Curses 程序	825
21.6 清空屏幕	828
21.7 决定终端大小	829
21.8 改变文本的颜色	831
21.9 读入密码	834
21.10 允许用 Readline 编辑输入	835
21.11 使得键盘指示灯闪烁	836
21.12 用 Tk 创建一个 GUI 应用程序	839
21.13 用 wxRuby 创建一个 GUI 应用程序	843
21.14 用 Ruby/GTK 创建一个 GUI 应用程序	847
21.15 用 RubyCocoa 创建一个 Mac OS X 应用程序	851
21.16 用 AppleScript 得到用户输入	859
第 22 章 用其他语言扩展 Ruby	861
22.1 为 Ruby 编写 C 扩展程序	862
22.2 在 Ruby 中使用 C 库	865
22.3 通过 SWIG 调用 C 库	869
22.4 通过 SWIG 调用 C 库	871
22.5 用 JRuby 使用 Java 库	874
第 23 章 系统管理	878
23.1 脚本化外部程序	879
23.2 管理 Windows 服务	880

23.3	作为另一个用户运行代码	882
23.4	不用 cron 或 at 运行周期性任务	884
23.5	删除匹配正则表达式的文件	885
23.6	批量重命名文件	888
23.7	找到复制的文件	891
23.8	自动备份	894
23.9	在用户目录下规范化所有权和权限	895
23.10	为给定用户杀死所有进程	898

前言

人生苦短

这是一本关于秘诀的书籍，其中包括：对一般问题的解决方案、复制和粘贴代码段、解释、示例和简短指南。

本书意欲为读者节省时间。人们总是说，时间就是金钱，但时间也是个人生命的构成。我们的生命应当花在创造新事物上，而不是用于抗击我们自己的错误或者解决那些已经解决过的问题。我们提供此书的愿望是：它的所有读者所节省的时间远远超过我们编写它所花掉的时间。

Ruby编程语言本身是一种非常节省时间的工具，与其他编程语言相比，它能产生更高的生产率，因为用户会花费更多的时间让计算机做自己想做的事情，而思考语言本身的时间则较少。但是，对于一名 Ruby 程序员，可能有很多情况即使没做任何事情也花费了很多时间，下面是我们遇到过的各个方面：

- 将时间花在编写通用算法的 Ruby 实现上。
- 将时间花在调试通用算法的 Ruby 实现上。
- 将时间花在发现和修正 Ruby 特有的缺陷上。
- 将时间花在应当自动执行的重复性任务上（包括重复性编程任务！）。
- 将时间花在重复其他人已经公开实现过的工作上。
- 将时间花在搜索运行 X 的库上。
- 将时间花在对运行 X 的多个库进行评估和确定上。