



Jakarta Commons Cookbook™ 中文版

Timothy M. O'Brien 著

乔健 张颖 译

倪森 审校

O'REILLY®



清华大学出版社

Jakarta Commons Cookbook™

中文版

Timothy M. O'Brien 著

乔健 张颖 译

倪森 审校

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Paris • Sebastopol • Taipei • Tokyo

O'Reilly Media, Inc. 授权清华大学出版社出版

清华大学出版社

图书在版编目 (CIP) 数据

Jakarta Commons Cookbook™ 中文版 / (美) 奥布莱恩 (O'Brien, T. M.) 著; 乔健等译. —北京: 清华大学出版社, 2006.11

书名原文: Jakarta Commons Cookbook

ISBN 7-302-13964-4

I. J… II. ①奥… ②乔… III. JAVA 语言—程序设计 IV. TP312

中国版本图书馆 CIP 数据核字 (2006) 第 120640 号

北京市版权局著作权合同登记

图字: 01-2006-5904 号

Copyright ©2004 by O'Reilly Media, Inc.

Authorized Simplified Chinese translation edition, by O'Reilly Media, Inc., is published by Tsinghua University Press, 2006. Authorized translation of the original English edition, 2004 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 2004 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 2006 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。

版权所有, 翻印必究。举报电话: 010-62782989 13501256678 13801310933

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。

本书防伪标签采用特殊防伪技术, 用户可通过在图案表面涂抹清水, 图案消失, 水干后图案复现; 或将表面膜揭下, 放在白纸上用彩笔涂抹, 图案在白纸上再现的方法识别真伪。

书 名 / Jakarta Commons Cookbook™ 中文版

书 号 / ISBN 7-302-13964-4/TP · 8397

责任编辑 / 常晓波

封面设计 / Emma Colby, 张健

出版发行 / 清华大学出版社 (www.tup.tsinghua.edu.cn)

地 址 / 北京清华大学学研大厦 (邮政编码 100084)

经 销 / 各地新华书店

印 刷 / 清华大学印刷厂

装 订 / 三河市李旗庄少明装订厂

开 本 / 178 毫米 × 233 毫米 16 开本 24.5 印张 531 千字

版 次 / 2006 年 11 月第 1 版 2006 年 11 月第 1 次印刷

印 数 / 1 ~ 3000 册

定 价 / 45.00 元 (册)

O'Reilly Media, Inc. 介绍

为了满足读者对网络 and 软件技术知识的迫切需求，世界著名计算机图书出版机构 O'Reilly Media, Inc. 授权清华大学出版社，翻译出版一批该公司久负盛名的英文经典技术专著。

O'Reilly Media, Inc. 是世界上在 Unix、X、Internet 和其他开放系统图书领域具有领导地位的出版公司，同时也是联机出版的先锋。

从最畅销的《The Whole Internet User's Guide & Catalog》（被纽约公共图书馆评为二十世纪最重要的 50 本书之一）到 GNN（最早的 Internet 门户和商业网站），再到 WebSite（第一个桌面 PC 的 Web 服务器软件），O'Reilly Media, Inc. 一直处于 Internet 发展的最前沿。

许多书店的反馈表明，O'Reilly Media, Inc. 是最稳定的计算机图书出版商——每一本书都一版再版。与大多数计算机图书出版商相比，O'Reilly Media, Inc. 具有深厚的计算机专业背景，这使得 O'Reilly Media, Inc. 形成了一个非常不同于其他出版商的出版方针。O'Reilly Media, Inc. 所有的编辑人员以前都是程序员，或者是顶尖级的技术专家。O'Reilly Media, Inc. 还有许多固定的作者群体——他们本身是相关领域的技术专家、咨询专家，而现在编写著作，O'Reilly Media, Inc. 依靠他们及时地推出图书。因为 O'Reilly Media, Inc. 紧密地与计算机业界联系着，所以 O'Reilly Media, Inc. 知道市场上真正需要什么图书。

目录

前言	1
第 1 章 扩展 J2SE 平台	11
1.0 简介	11
1.1 获取 Commons Lang	12
1.2 加入 Commons-User 邮件列表	13
1.3 获取 Commons Lang 源代码	14
1.4 自动生成 toString() 内容	15
1.5 自定义 toString() 内容	17
1.6 自动生成 hashCode() 和 equals()	19
1.7 自动实现 compareTo()	21
1.8 输出数组内容	23
1.9 克隆和反转数组	24
1.10 对象数组和基本类型数组的互相转换	27
1.11 在数组中搜索特定项	29
1.12 由多维数组生成 Map	30
1.13 格式化日期	32
1.14 舍入 Date 对象	34

1.15 截取 Date 对象	36
1.16 创建枚举量	37
1.17 生成唯一的数字标识符	41
1.18 验证方法参数	44
1.19 度量时间	46
第 2 章 处理文本	49
2.0 简介	49
2.1 获取 StringUtils 和 WordUtils	50
2.2 检查空字符串	50
2.3 缩减字符串	52
2.4 分割字符串	54
2.5 搜寻嵌套字符串	55
2.6 裁减字符串	57
2.7 字符串的 chomp 处理	58
2.8 创建醒目的标题	59
2.9 反转字符串	59
2.10 单词折回	61
2.11 检测字符串内容	62
2.12 检测字符串出现频率	64
2.13 解析格式化字符串	66
2.14 计算字符串的差别	68
2.15 使用 Commons Codec	70
2.16 获取 Commons Codec 源代码	71
2.17 计算 Soundex 编码	72
第 3 章 JavaBeans	74
3.0 简介	74
3.1 以图形方式表示 Beans	75
3.2 获取 Commons BeanUtils	78

3.3	获取 Commons BeanUtils 源代码	79
3.4	访问简单的 Bean 属性	80
3.5	访问嵌套 Bean 属性	82
3.6	访问带有下标的 Bean 属性	84
3.7	访问 Map 型 Bean 属性内容	85
3.8	访问一个嵌套的带有下标的 Map 型简单 Bean 属性	87
3.9	确定 Bean 属性的类型	88
3.10	比较 Beans	90
3.11	复制 Bean 属性	93
3.12	Bean 的克隆	95
3.13	设置 Bean 属性	95
3.14	检查可否访问属性	97
3.15	通过 Predicates 验证 Beans	98
3.16	创建包含 Bean 属性的 Map	100
3.17	使用 Map 封装 Bean	101
3.18	创建动态 Bean	103
3.19	以字符串形式读取和设置属性	105

第 4 章 函子 107

4.0	简介	107
4.1	获取 Commons Collections	108
4.2	获取 Commons Collections 源代码	109
4.3	Comparator 的功能反转	109
4.4	Comparators 的链接	112
4.5	对 null 的比较	114
4.6	以固定顺序比较	116
4.7	使用简单的 Predicates	119
4.8	自定义 Predicate	124
4.9	创建复合 Predicate	126
4.10	对象转换	130

4.11 创建转换链	131
4.12 实现条件转换	133
4.13 编写 Closure	135
4.14 链接多个 Closure	136
4.15 使用 Closures 模拟条件语句	137
4.16 使用 Closure 模拟循环	140

第 5 章 Collections 142

5.0 简介	142
5.1 获取 Commons Collections	142
5.2 使用循环迭代器	143
5.3 遍历 ArrayList	145
5.4 使用 Predicate 筛选 Collection	147
5.5 遍历相异的元素	150
5.6 使用 Bag	151
5.7 使用 Buffer	156
5.8 创建优先队列	158
5.9 使用阻塞式缓冲区	162
5.10 在 Map 中存储多个值	165
5.11 根据值检索相应的键	167
5.12 使用大小写不敏感的 Map	170
5.13 创建特定类型的 Collection 和 Map	172
5.14 限制 Map 的取值	176
5.15 限制 List 的内容	179
5.16 转换 Collections	180
5.17 创建 LRU 缓存	183
5.18 使用 LazyMap	184
5.19 统计 Collection 中对象的出现次数	188
5.20 执行集合操作	190
5.21 不经强制转换检索 Map 的值	192

第 6 章 XML	194
6.0 简介	194
6.1 获取 Jakarta Commons Digester	195
6.2 将 XML 文档转为对象	196
6.3 带命名空间的解析	200
6.4 创建简单 XML 命令语言	204
6.5 变量置换和 XML 解析	207
6.6 获取 Jakarta Commons Betwixt	210
6.7 转换 Beans 为 XML 文档	211
6.8 定制由对象生成的 XML	213
6.9 转换 XML 文档为 Beans	216
第 7 章 应用程序基础设施	219
7.0 简介	219
7.1 获取 Commons CLI	219
7.2 解析简单命令行	220
7.3 解析复杂命令行	222
7.4 显示使用方法信息	223
7.5 获取 Commons Configuration	225
7.6 通过属性文件配置应用程序	226
7.7 通过 XML 配置应用程序	227
7.8 使用复合配置	229
7.9 获取 Commons Logging	232
7.10 使用抽象日志接口	233
7.11 指定日志的实现	234
7.12 获取 Apache Log4J	235
7.13 根据属性文件配置 Log4J	236
7.14 通过 XML 配置 Log4J	238

第 8 章 Math	240
8.0 简介	240
8.1 使用分数	241
8.2 在数组里寻找最大和最小数	243
8.3 处理数字的范围	244
8.4 生成随机变量	246
8.5 获取 Commons Math	247
8.6 计算简单的单变量统计量	248
8.7 解线性方程系统	250
8.8 带复数的算术操作	252
8.9 在变量间建立关联	254
8.10 估计剩余处理时间	256
 第 9 章 模板	 261
9.0 简介	261
9.1 获取 Commons JEXL	264
9.2 使用表达式语言	265
9.3 在表达式中调用方法	268
9.4 以表达式语言外置逻辑	269
9.5 获取 Jakarta Velocity	271
9.6 使用简单模板语言	272
9.7 编写含有条件和循环的模板	274
9.8 在模板引擎中使用宏	277
9.9 在模板中调用方法	280
9.10 获取 FreeMarker	284
9.11 使用复杂的脚本引擎	285
9.12 从模板引擎访问 XML 文档	288
9.13 在 Web 应用程序中使用 Velocity	291
9.14 在 Web 应用程序中使用 FreeMarker	292
9.15 在 Eclipse 中编写模板	294

第 10 章 I/O 与网络	295
10.0 简介	295
10.1 获取 Commons IO	296
10.2 拷贝 Stream、byte[] 数组、Reader 和 Writer	297
10.3 关闭 Stream、Reader 和 Writer	298
10.4 输出易于理解的文件大小值	300
10.5 拷贝文件、字符串和 URL	302
10.6 递归删除目录	303
10.7 获取目录大小	304
10.8 文件 touch 操作	304
10.9 筛选文件	305
10.10 测量 Stream 流量	306
10.11 拆分 OutputStream	308
10.12 获取 Jakarta ORO	308
10.13 使用 glob 表达式和 Perl5 正则表达式列举文件	309
10.14 获取 Commons Net	310
10.15 编写 FTP 客户端	311
10.16 通过 SMTP 发送邮件	313
10.17 查看 POP3 邮箱	315
 第 11 章 HTTP 和 WebDAV	 318
11.0 简介	318
11.1 获取 Jakarta HttpClient	318
11.2 获取 Jakarta HttpClient 源代码	320
11.3 执行 HTTP GET	321
11.4 在查询字符串中发送参数	323
11.5 通过条件 GET 检索内容	325
11.6 调试 HTTP	328
11.7 执行 HTTP POST 请求	330
11.8 发送来自文件的 POST 数据	332

11.9 通过 Multipart POST 上传文件	333
11.10 Basic 型身份验证	335
11.11 NTLM 身份验证	338
11.12 处理 Cookie	339
11.13 处理重定向	342
11.14 SSL	345
11.15 使用自签发的证书	346
11.16 获取 Jakarta Slide	347
11.17 连接 WebDAV 资源	349
11.18 修改 WebDAV 资源	351
第 12 章 搜索和筛选	353
12.0 简介	353
12.1 获取 Commons XPath	354
12.2 通过 XPath 查询对象图表	354
12.3 搜索一组简单对象	357
12.4 在复杂的对象图表中使用 XPath 查询	359
12.5 获取 Jakarta Lucene	362
12.6 为 XML 文档创建索引	363
12.7 在文档索引中搜索指定词语	368
12.8 找出索引中词语出现的频率	372

前言

在本书中，你能找到一组由 Jakarta Commons (Commons, <http://jakarta.apache.org/commons>) 是一组小巧而流行的组件，位于 Apache Software Foundation 的 Jakarta 项目下。它囊括从简单到复杂的诸多功能，很多开发者认为其中的某些库对 Java 项目而言是不可或缺的。这些组件获得了广泛使用，有可能已经位于你的类路径中了。如果你在开发涉及 Jakarta Struts、Apache Tomcat 或 Hibernate 的应用程序，那你的类路径中就有 Commons BeanUtils 和 Commons Logging。如果刚以默认配置方式安装了 Red Hat Enterprise Linux，那么可以在 `/usr` 目录里的某个位置找到 Commons 库。如果下载 Sun's J2EE 1.4 SDK，你也将同时得到了 Commons Logging 和 Commons Launcher。虽然 Jakarta Commons 几乎无处不在，可还有很多人不了解这些组件的强大功能。

本书使用 Jakarta Commons 处理具体的实现细节，并回答了这样的问题：如何解析 XML？如何串行化 beans？有没有处理 Collection 的更好方法？如何处理 HTTP 并跟踪 cookie？这些具体实现在面对全局设计时往往被忽视。考虑拥有一个设计良好的可靠架构的复杂 J2EE 系统。从居高临下的角度看来，其全局（高层）设计显得非常合理，不过如果涉足其细节，你就会发现每个 Servlet 或 JSP 都充斥着难以维护又无必要的代码。究其原因，主要在于开发者不知道使用一些省时省力的工具，譬如 BeanUtils、Collections 或 Digester。更糟的情况是，开发者可能花了一周时间实现了 Commons BeanUtils 提供的大部分功能，最后却发现 BeanUtils 早已在项目的类路径中了。综上所述，了解 Jakarta Commons 提供的内容将有助于开发者在底层开发时做出正确的决定。

没人会傻到去写自己的 XML 解析器，但是常有开发者编写的定制组件功能雷同于已有的免费组件。以一组静态实用方法为例，这些方法几乎出现在每个复杂项目中。譬如，把文件内容读到字符串的通用过程能被重构到 CommonFileUtils 类中，而把 DOM Document 对象转为 bean 的功能则可通过自定义代码中的一组类予以实现。Jakarta

Commons为这些问题都提供了解决方案。阅读本书将有助于读者避免这些不必要的重复实现。

很多开发者对这些组件有一些大概了解,但是很少有人能花上数月或数星期的时间坐下来好好研读一下各组件相关的Tutorial、FAQ、blog以及邮件列表存档。由于要跟踪一系列开源社区的更新,上述工作量不可小觑。这也是为什么本书要将大部分内容组织成易于阅读和理解的章节。这些章节提供了快速上手Commons所需的知识,除此之外,如果需要更多信息,各小节的讨论和参考部分还为深入理解每种Commons组件内部实现提供了途径。

本书介绍的工具将使你节约大量时间,并能在你尚未意识到时提供一组替代选择。真希望自己能在五年前读到这样一本书,那会节省我很多学习时间,也能帮我避免过去一些代价昂贵的设计失误。请读者按照自己的情况阅读本书。如果只对Commons Collections感兴趣,那么浏览第5章内容就能使你快速了解Collections。如果希望了解Jakarta Commons中各主要项目的概况,就应该从头至尾通读本书。本书通过一种半引用半散文的结构,使读者能够定制自己的阅读方式。希望本书读起来和编写时一样有趣。

本书的内容

本书内容包括Jakarta Commons组件、部分Jakarta项目以及一个位于Apache Software Foundation之外的项目。本书涉及到下列组件:

- Jakarta Commons BeanUtils
- Jakarta Commons Betwixt
- Jakarta Commons CLI
- Jakarta Commons Codec
- Jakarta Commons Collections
- Jakarta Commons Configuration
- Jakarta Commons Digester
- Jakarta Commons HttpClient
- Jakarta Commons ID
- Jakarta Commons IO
- Jakarta Commons JEXL

- Jakarta Commons JXPath
- Jakarta Commons Lang
- Jakarta Commons Logging
- Jakarta Commons Math
- Jakarta Commons Net
- Apache Log4J
- Jakarta Velocity
- FreeMarker
- Jakarta Lucene
- Jakarta Slide

所有这些组件都将在后面各章予以详细介绍。这里列举一下各章主要内容：

第1章, *Java2* 平台的扩展

这一章介绍 Commons Lang 组件, 主要讨论如何自动生成 toString()、处理数组、格式化及舍入日期值、处理枚举量、生成标识符、以及测量时间值等内容。这一章还涉及带 Commons ID 的全局唯一标识符的生成问题。

第2章, 处理文本

与脚本语言 (如 Perl) 不同, Java 缺乏扩展文本处理能力, 而 Commons Lang 的 StringUtils 则提供了一些用于处理文本的实用方法。这一章内容涉及 StringUtils、WordUtils 和 Commons Codec。

第3章, *JavaBeans*

在 Java 中到处能看到 Bean, 从 Jakarta Struts 到 Hibernate 都是如此。Bean 是对象模型中的信息单元。这一章将介绍 Commons BeanUtils, 它是 Jakarta Commons 中最广为使用的组件之一。

第4章, 函子

函子是一种程序基本设计思想, 它把程序抽象为一组功能对象的集合。Commons Collections 引入了 Predicate、Transformer 与 Closure, 还有用于对控制结构与循环进行建模的各种函子。这一章将说明如何把函子运用到程序中。

第5章, 集合类

这一章介绍了迭代器、条件筛选、缓冲区、队列、双向 Map、类型安全集合、受限集、延迟存取 Map 和集合操作等内容。本章涉及 Commons Collections, 新引入的集合类型以及针对多种集合类的函子应用程序。

第6章, XML

如果常要解析或创建XML文档,这一章内容将为你提供一些替代标准解析器API(如SAX、DOM和JDOM)的途径。本章将引入 Commons Digester、 Commons Betwixt以及 Commons JXPath。

第7章, 应用程序基础设施

这一章将首先介绍 Commons Configuration, 它能解析属性文件与XML配置文件。本章其他内容则将阐述如何使用 Commons CLI 解析一组由必选项和可选项组成的复杂命令行选项。本章还将详细介绍如何配置使用 Commons Logging 和 Apache Log4J。

第8章, Math

这一章将关注 Commons Lang 和 Commons Math 包中的简单数学功能。本章介绍的类能处理分数、复数、矩阵以及简单的单变量统计。

第9章, 模板

这章内容将涉及从简单表达式语言(如 Commons JEXL)到复杂模板引擎(如 Jakarta Velocity 与 FreeMarker)的诸多内容。本章还将演示如何把 Velocity 与 FreeMarker 集成到 Jakarta Tomcat 这样的 J2EE Servlet 容器中。

第10章, I/O 和网络

这一章将介绍 Commons IO 和 Commons Net 包。前者含有一系列处理流和文件的实用程序,而后者包括有面向 FTP、POP 和 SMTP 协议的简单客户端。

第11章, HTTP 和 WebDAV

如果你需要通过 HTTP 进行通信,请阅读这一章。本章将介绍 Jakarta HttpClient, 以及来自 Jakarta Slide 的 WebDAV 客户端库。

第12章, 搜索和筛选

Commons JXPath 可用于通过 XPath 表达式访问集合类与对象图表。而 Jakarta Lucene 则是一个全功能的搜索引擎,可索引任何结构的文档。本章将介绍如何搭配使用 Lucene 和 Commons Digester。

由于时间和篇幅所限,我不得不只把部分项目列入本书。像 Velocity、FreeMarker 和 Log4J 这样的非 Commons 组件也能成为本书的一部分,是因为它们符合简单且易复用的组件标准。其他一些组件未能列入本书,是因为在本书编撰时它们还在开发中,或者无法用简短的一节内容完成对它们的介绍。像 Commons DbUtils、DBCP、Discovery、Jelly、Launcher、Modeler、Pool、Primitives、Chain 以及 promising sandbox 这样每个组件的内容都能填满一整卷书。有些项目,譬如 Jakarta HiveMind, 起源于 Commons Sandbox 的子项目,却直接作为 Jakarta 项目的子项目得以发展。Jakarta 中项目和组件的

分类也显得有点混乱。Jakarta ORO与Jakarta RegExp看上去很可能是Jakarta Commons的一部分，但却都是Jakarta的子项目。其他一些项目如Jakarta Commons HttpClient，则在近期升格为Jakarta的子项目而完全脱离了Commons。为免困惑，建议你将本书看作是一本介绍了Jakarta Commons及其他一些项目的图书。

为一系列更新迅速的组件写书的过程让我回忆起一种名为whack-a-mole的游戏。就当你刚为某个新版本更新完内容时，另一种组件又出现了。平均每一至二个星期就有一种Commons组件得以发布，所以在本书出现在书架上时，书中一些组件的版本已显陈旧。大体上Jakarta Commons能支持向后兼容性并保有一组稳定的公共接口。你学到的有关Commons BeanUtils 1.6的知识在Commons BeanUtils 1.7上还会继续适用。如果你发现某个组件发布了更新版本，请下载该新版本，并在O'Reilly站点上查看本书是否有相应的更新。

本书约定

你应该了解本书使用的一些约定。例如，菜单项之间将由_符号分隔，正如这里所示：File_New_Project。为了起强调作用，代码中新出现的各行将以高亮方式显示。示例代码将常常在脱离上下文的环境下提供；本书将不给出完整的类，而是仅列出与之相关的代码块。大部分示例将包含对Commons相关类的import语句，而其他import语句将被忽略。若有代码被省略或隐藏，将在其位置显示省略号：

```
import org.apache.commons.digester.Digester;
...
Digester digester = new Digester();
digester.doSomething();
```

此外，本书还使用了以下排版约定：

斜体字 (*Italic*)

表示新出现的名词、URL、电子邮件地址、文件名、文件扩展名、路径、目录以及Unix实用程序。

等宽字体 (Constant Width)

表示命令、选项、分支、变量、类型、类、命名空间、方法、模块、属性、参数、值、对象、时间、时间处理程序和XML标记。

等宽斜体 (*Constant Width Italic*)

显示应被用户提供的值替换的文本。