

O'REILLY®

# GOOGLE HACKS™

第二版  
涵盖 Gmail

巧妙使用网络搜索的技巧和工具



*Tara Calishain & Rael Dornfest* 著

*Craig Silverstein* 序

卞军 谢伟华 朱炜 译



电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY  
<http://www.phei.com.cn>

## GOOGLE HACKS 巧妙使用网络搜索的技巧和工具



对于Google，有很多东西是一般用户所不了解的。看起来简单、明了的Google界面掩饰了它下面深藏的Google索引、引擎的源动力和搜索语法的灵活性。Google Hacks义无反顾地潜入到深层的Google世界中，最大程度地挖掘了Google的潜能：挖掘信息而不仅是搜索，执行有效完成工作的任务，使用Google API自动执行复杂或者重复性的工作，想尽办法挖掘出每个搜索结果的价值。

考虑到超过80亿个网页，Google索引的规模已经是“Google Hacks”第一版出版时的四倍了。既然Google发展得这么快，本书也同样要发展，以收录进新的Google服务和用于原有功能的新hack程序。除了将过去的内容完全更新和重组外，这本“Google Hacks”还将继续给用户提供了很多令人叫绝的Google技巧。

- 通过Google的更具灵活性的搜索语法进行更巧妙的搜索；
- 使用Google Web API和一点Perl、PHP、Python、Java或.NET程序找到那些很难到达的页面的结果；
- 用Google Desktop将Google搜索的强大能量和灵活性放到你的桌面上，或者用Google Site Search将搜索功能放到你的站点中；
- 用Google AdWords穿越在线广告的屏障——使用站点中的AdSense来为自己赚钱；
- 清楚地了解Google如何检索用户的网站，了解更多关于Google的著名PageRank算法的内容；
- 让Google独特的、流行的在线邮件系统——Gmail来改变你使用电子邮件的方式。

“自从Google Hacks一出现，网络搜索，如果可能，就已经成为不可或缺的部分了……还能提供人们更需要的自动化作业以及在信息海洋中给一根“信息小针”定位的功能，而且，还在继续发展。”

——Craig Silverstein, Director of Techology, Google

图书分类：网络技术 > 搜索引擎

ISBN 7-121-01451-3



9 787121 014512 >



www.phei.com.cn

ISBN 7-121-01451-3

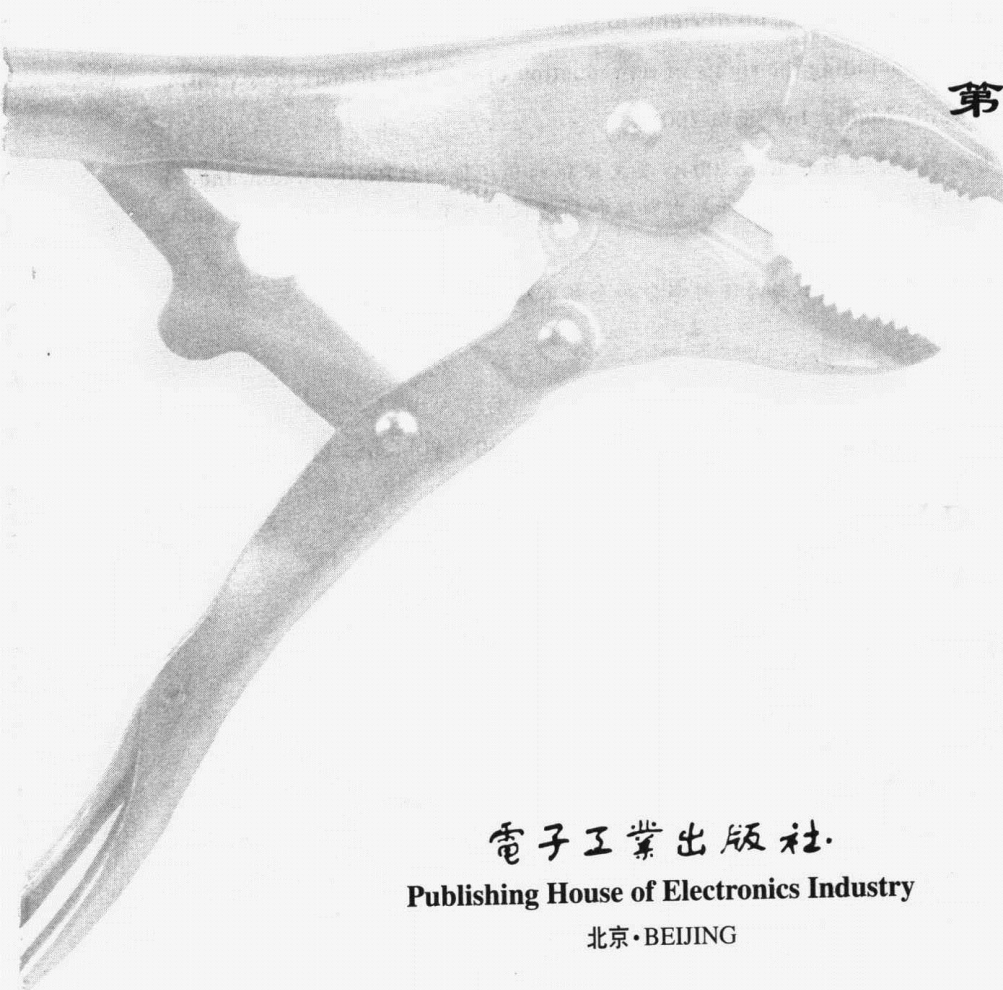
定价：49.00元

O'REILLY®

# GOOGLE HACKS™

巧妙使用网络搜索的技巧和工具

第二版



電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

## 图书在版编目 (CIP) 数据

Google Hacks 巧妙使用网络搜索的技巧和工具 (第二版) / 加利吉安 (Calishain, T.) 等著; 卞军, 谢伟华, 朱炜译. —北京: 电子工业出版社, 2006.1  
书名原文: Google Hacks Tips & Tools for Smarter Searching, Second Edition  
ISBN 7-121-01451-3

I. G... II. ①加... ②卞... ③谢... ④朱... III. 因特网—情报检索—基本知识  
IV. G252.7

中国版本图书馆 CIP 数据核字 (2005) 第 067332 号

版权贸易合同登记号

图字: 01-2004-1767 号

©2004 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Publishing House of Electronics Industry, 2004. Authorized translation of the English edition, 2003 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.

英文原版由 O'Reilly Media, Inc. 出版 2004。

简体中文版由电子工业出版社出版 2004。英文原版的翻译得到 O'Reilly Media, Inc. 的授权。此简体中文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, 未得书面许可, 本书的任何部分和全部不得以任何形式重制。

书 名 / Google Hacks 巧妙使用网络搜索的技巧和工具 (第二版)

书 号 / ISBN 7-121-01451-3

责任编辑 / 顾慧芳

封面设计 / Edie Freedman, 张健

出版发行 / 电子工业出版社 (<http://www.phei.com.cn>)

地 址 / 北京市海淀区万寿路 173 信箱 (邮政编码 100036)

经 销 / 全国新华书店

印 刷 / 北京市天竺颖华印刷厂

开 本 / 787 毫米 × 980 毫米 16 开本 32 印张 350 千字

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

印 数 / 0001-4000 册

定 价 / 49.00 元 (册)

---

# 目录

序 .....	1
第一版的序 .....	3
致谢 .....	5
前言 .....	15
<b>第一章 网页 .....</b>	<b>27</b>
引言: Google 网络搜索的基本知识 .....	27
全字通配符 .....	32
10 个单词的限制 .....	34
特殊的语法结构 .....	35
混合语法结构 .....	39
高级搜索 .....	45
快速链接 .....	47
语言工具 .....	50
分析搜索结果 .....	53
设置参数 .....	56
弄清楚 Google 的 URLs .....	58
1. 浏览 Google 的地址名录 .....	60
2. 及时从 Google 上得到初步印象 .....	63

3. Google 结果随时间变化的图表 .....	66
4. 显示 Google 搜索结果 .....	68
5. 检查拼写 .....	73
6. Google 电话簿：让 Google 的手指为你指路 .....	75
7. 全球化的视角与 Google 本地查询 .....	78
8. 跟踪股票信息 .....	83
9. 搜索辞典 .....	85
10. 查找定义 .....	87
11. 搜索归档文件 .....	93
12. 查找信息所在的目录 .....	95
13. 搜索 Weblog 的注释 .....	97
14. 搜索所有基本的关键词 .....	99
15. 重复使用关键词 .....	102
16. 查询特定的日期范围 .....	103
17. 计算 Google 上的 Centuryshare .....	107
18. Google 搜索表单妙用 .....	110
19. 超越 Google 的高级搜索 .....	113
20. 使用 Google 工具进行翻译 .....	115
<b>第二章 高级页面 .....</b>	<b>120</b>
引言：一些假设 .....	121
21. 同义词 .....	122
22. 在 Google Box 中获得 Google 搜索结果 .....	125
23. 构造 Google 网页目录的 URL .....	131
24. 寻找菜谱 .....	133
25. 随时跟踪搜索结果的数量 .....	137
26. 手气真不错 (Feeling Really Lucky) .....	141
27. (有意地) 获取随机结果 .....	144
28. 改变查询序列的组合方式 .....	146
29. 衡量查询关键词的权重 .....	151
30. 获取顶级目录下的搜索结果 .....	152

31. 搜索特殊字符 .....	158
32. 深入挖掘网站层级目录 .....	161
33. 分析搜索结果的域频率 .....	164
34. 测量 Google “记忆占有率” .....	168
35. SafeSearch 验证 URL .....	170
36. 搜索 Google 主题 .....	174
37. 查找最大的网页 .....	178
38. 进行邻接搜索 .....	180
39. 漫游 Google 邻域 .....	183
40. 运行 Google 流行度排行榜 .....	191
41. 从 Yahoo! Buzz 剪辑 Google 搜索条件 .....	199
42. 比较 Google 与其他搜索引擎的搜索结果 .....	203
43. 在 Yahoo!和 Google 上进行多点搜索 .....	206
44. Yahoo!的目录在 Google 中的记忆占有率 .....	209
45. 收集 Weblog —— 免费的 Google 搜索结果 .....	214
46. 用 Geotargeting 来预测趋势 .....	217
47. 在命令行调用 Google 计算器 .....	224
48. 构造一个可定制日期范围的搜索表单 .....	225
49. 搜索前一天的索引 .....	228
<b>第三章 图像 .....</b>	<b>232</b>
引言: Google 图像高级搜索界面 .....	233
Google 图像专用语法 .....	234
50. 移用某个公司或者产品的标志 .....	235
51. 浏览世界各地的照相簿 .....	237
52. Google 绘图法: 你身边的街头艺术 .....	240
53. 抓取地图 .....	244
<b>第四章 新闻组和论坛 .....</b>	<b>247</b>
引言: Google 新闻组 .....	247
Google 新闻组的搜索语法 .....	248

高级新闻组搜索 .....	250
充分利用 Google 新闻组 .....	251
接收 Google 新闻警报 .....	251
Google 以外的新闻搜索引擎 .....	251
54. 剪辑 Google 新闻组 .....	252
55. 可视化 Google 新闻组 .....	256
<b>引言: Google 论坛 .....</b>	<b>258</b>
十大层级 .....	258
浏览讨论组 .....	259
Google 论坛专用语法 .....	260
高级论坛搜索 .....	262
56. 通过 Google 论坛二深入地了解 Google 论坛 .....	263
57. 剪辑 Google 论坛 .....	268
58. 简化 Google 论坛的 URL .....	272

## **第五章 附件 .....** 274

59. 通过 Google 警报记录用户的搜索 .....	274
60. 将 Google 放到工具栏或者桌面上 .....	278
61. “Google” 你的桌面 .....	284
62. 使用 Bookmarklet 的 Google .....	293
63. “Google” Word 文档 .....	295
64. 通过电子邮件访问 Google .....	304
65. 使用即时通信进行 Google 搜索 .....	306
66. “Google” 用户的 IRC (在线聊天系统) .....	309
67. 外出时如何使用 Google .....	313
68. Google 实验室 .....	321
69. 看看 Google 对___的看法 .....	322
70. 搜索引擎腰带扣 .....	325



<b>第六章 Gmail .....</b>	<b>336</b>
引言: Gmail 搜索语法 .....	337
额外的资源 .....	339
71. 收集 Gmail 邀请 .....	340
72. 建立和使用习惯的地址 .....	342
73. 将地址簿导入 Gmail .....	344
74. 将邮件导入到 Gmail .....	354
75. 导出 Gmail 邮件 .....	362
76. 漫步 Lighter Side .....	364
77. 在户外使用 Gmail .....	369
78. 使用 Gmail 作为 Linux 文件系统 .....	375
79. 将 Gmail 作为 Windows 驱动器使用 .....	380
80. Gmail 编程 .....	382
<b>第七章 广告 .....</b>	<b>386</b>
引言: Google AdSense .....	387
Google AdWord .....	387
81. 充分利用 AdWords .....	387
82. 生成 Google 的 AdWords .....	392
83. 剪辑 Google Adwords .....	394
84. 衡量 AdWords 关键词的价值 .....	397
85. 提供备用广告 .....	409
<b>第八章 Web 管理 .....</b>	<b>414</b>
引言: Google 对于 Web 管理员的重要性 .....	414
神秘的 PageRank .....	415
机会均等的神秘算法 .....	415
紧跟 Google 的变化 .....	415
一个词: 放松 .....	416
86. 向 Web 管理员介绍 Google .....	416
87. 解密 PageRank 的算法 .....	420

88. 达到每天 15000 次的访问量的 26 个步骤 .....	423
89. 进驻搜索引擎 .....	429
90. 为 Google 的来访做好清理准备工作 .....	433
91. 从 Google 清除你的材料 .....	435
<b>第九章 Google 编程 .....</b>	<b>439</b>
引言: 注册申请与 Google 服务条款 .....	440
Google Web API 的开发者工具包 .....	441
在程序中使用许可证密钥 .....	441
什么是 WSDL? .....	442
深入理解 Google API 的查询方式 .....	443
深入理解 Google API 的响应 .....	447
对 Spidering 和 Scraping 的解释 .....	450
92. Google Web API 的 Perl 编程 .....	452
93. 安装 SOAP::Lite Perl 模块 .....	454
94. Google 编程的 Net::Google Perl 模式 .....	457
95. 循环读取 10 条结果 .....	459
96. Google 的 PHP 编程 .....	462
97. Google 的 Java 编程 .....	464
98. Google 的 Python 编程 .....	467
99. Google 的 C# 与 .NET 编程 .....	470
100. 使用 VB.NET 进行 Google 编程 .....	473
<b>索引 .....</b>	<b>477</b>

---

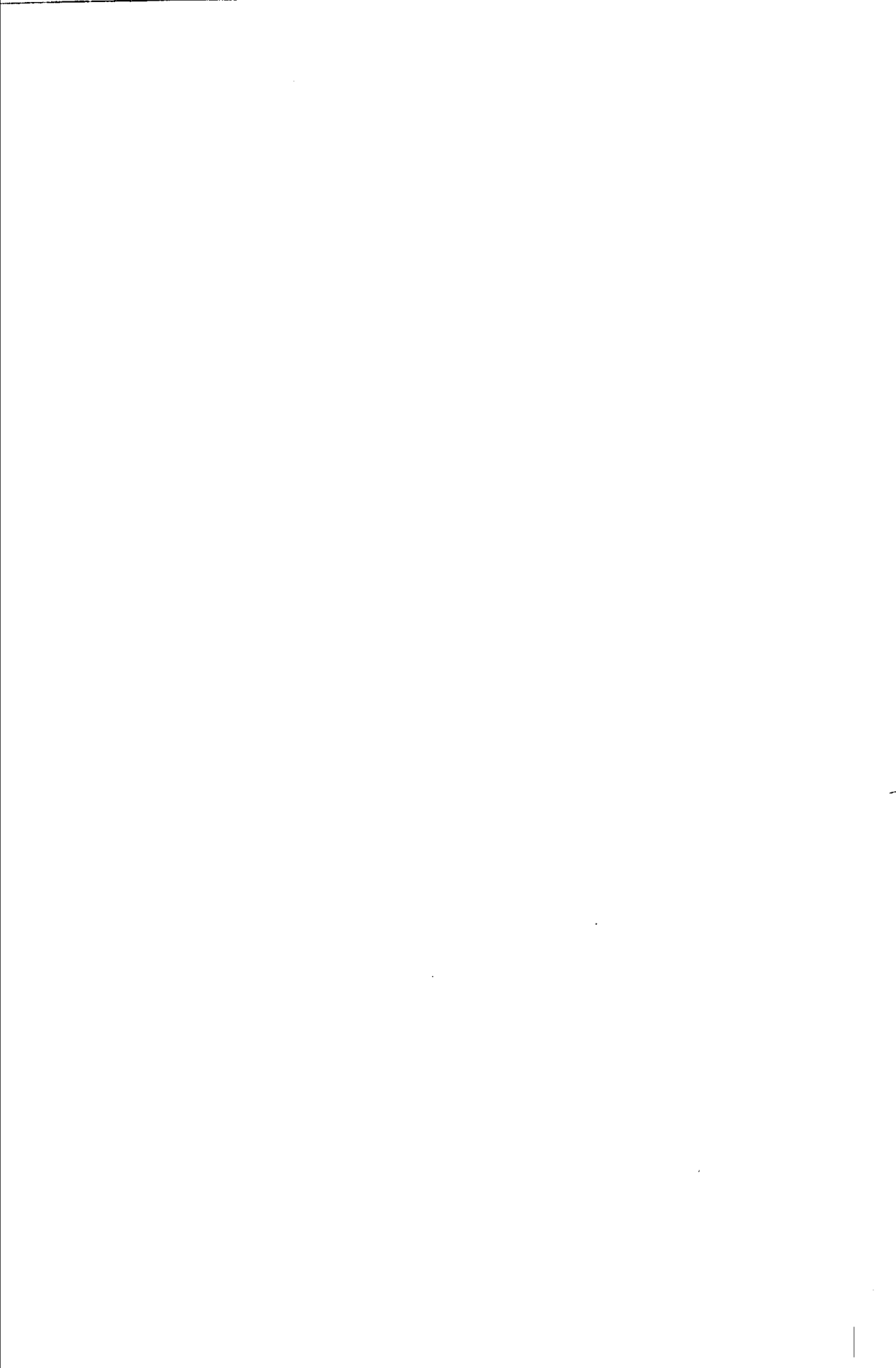
# 序

第一版“Google Hack”问世的时候，我们感到很惊奇，居然有这么多*hacking*值得写，而且能写成一本书。不过现在已经不再觉得奇怪了。人们一直在给Google的无数功能寻找着比我们想像中更多的改变、调整甚至是随意浏览的方法。

在“Google Hack”问世之后的18个月内，网络搜索，如果可能的话，就已经成为不可或缺的部分了。它不仅给用户提供更多的信息——通过电子邮件、电脑硬盘，以及从前书面内容的新的数字化版本——还能提供人们更需要的自动化作业，以及在信息的海洋中给一根“信息小针”定位的功能，而且还在继续发展。

我们希望用户可以尽情浏览这本新的“Google Hack”，继续帮助我们让更多的全球信息在网络上传播和使用。

——Google 技术董事 Craig Silverstein



---

# 第一版的序

我们最初使用 Google 时，未曾料想到它会变得如此重要。搜索引擎有朝一日会成为网络发展的催化剂，这曾经是个遥远的梦想。人们对 Google 日益增长的兴趣使我们备感荣幸，本书对 Google 搜索技术做了最全面、最详尽的阐述，在它即将付梓之际，我们要衷心感谢本书的作者。

搜索是一个令人惊奇的研究领域，因为它为我们向大家提供有效信息提供了无限的可能性。我们与本书的作者一起，鼓励读者利用本书来发现和创造新的搜索方法。Google 的任务是将世界各地的信息组织起来，并使其得以广泛地发布和应用。我们也欢迎你对此提出建议。

黑客精神 (hacking) 具有强大的创造性，为网络的发展提供了源动力。我们作为软件开发者，对这本书的创新思想表示赞同。我们也喜欢创新，因此很高兴地发现，自己闲暇时在 Google 上做过的许多试验在本书中也有精彩的描述。

为了适应互联网一日千里与瞬息万变的本质，Google 也在不断地改变自己的算法。阅读本书时，请注意，书中的例子在今天虽然有效，但可能转眼间就会成为昨日黄花，因为 Google 自身还在不断革新与完善。欢迎你紧随发展潮流，并积极参与到当前关于搜索技术的讨论中来。

事实上，Google 的每一位工程师都曾使用过 O'Reilly 的出版物来辅助他们的工作。O'Reilly 的书是 Google 工程图书馆中的主要资料。我们希望本书对于读者也如同 O'Reilly 出版物对于 Google 一样的有用。

作为世界上网络文档的最大收集者, Google同时反映了网络的状况。本书中黑客精神不仅用于 Google 而且也反映了当今和未来网络所能释放的巨大潜力。对于搜索爱好者来说, 本书无疑是个巨大的资源, 我们希望你和我们一起喜欢它。

谢谢!

—— Google 工程组

2002 年 12 月 11 日

美国加利福尼亚州, 山景城

---

# 致谢

## 作者简介

Tara Calishain 是网络搜索及在线信息资源每周时事通讯“ResearchBuzz” ([www.researchbuzz.com](http://www.researchbuzz.com)) 的编辑。也是“Searcher”杂志的专栏作家。她从1996年起就开始撰写有关搜索引擎和搜索的内容；她近期撰写的书还包括“Web Search Garage”。

Rael Dornfest 是 O'Reilly Media 的主要技术官员。他服务于 O'Reilly 网络和 O'Reilly 出版物，负责评测、试验、编程、协调、研究与写作。他是 O'Reilly Hack 系列丛书 (<http://hacks.oreilly.com>) 的系列编辑，还编写、合作编写以及辅助编写了多部 O'Reilly 著作，包括“Mac OS X Panther Hacks”，“Mac OS X Hacks”，“Google Pocket Guide”，“Google: The Missing Manual, Essential Blogging”和“Peer to Peer: Harnessing the Power of Disruptive of Technology”。他也是 O'Reilly Emerging 技术会议 (<http://conferences.oreilly.com/etech>) 的召集人。Rael 还利用闲暇时间开发自由软件，尤其是 Bloxom weblog 应用程序 (<http://www.bloxom.com>)，还任“MobileWhack” (<http://www.mobilewhack.com>) 的主编，他还时常维护他的 Raelity Bytes Weblog ([www.raelity.org](http://www.raelity.org))。

## 供稿人

下列人士为本书在构思、技术与写作上提供了帮助：

- Tim Allwine 是 O'Reilly Media 的高级软件工程师。他为市场研究工作

组开发软件 —— 开发各种用于从互联网上收集数据的剪辑工具 —— 并参与开发 O'Reilly 的 Web 服务。

- DJ Adams (<http://www.pipetree.com/qmacro>) 是一个 SAP 黑客，他用几天的时间开发了工作控制语言和 S/370 汇编程序，踩着他的溜冰板逛了伦敦中心。现在，他正在研究 NetWeaver 技术，并把剩下的精力都用在了探讨 REST, RDF 和 Jabber。他编写了 O'Reilly 公司的“Programming Jabber: Extending XML Messaging”，还合作编写了 O'Reilly 的“Google Pocket Guide”。他现在和 Sabine、Joseph 一起住在欧洲。
- AvaQuest (<http://www.avaquest.com/>) 是一家坐落在马萨诸塞州的 IT 服务公司。他们专门研究利用高级的信息检索、分类与文本挖掘技术解决实际问题。Google Web API 发布后不久，2002 年夏天，AvaQuest 公司的顾问 Nathan Treloar、Sally Kleinfeldt 与 Peter Richards 即从他们正在开发的一个 Web 数据挖掘咨询项目中成功地推出了 GooglePeople 和 GoogleMovies。
- Paul Bausch (<http://www.onfocus.com/>) 是一位来自俄勒冈州的 Web 开发者与自由撰稿人。他是 Blogger Weblog 软件的开发者之一。最近，他正在与人合写一本关于 Weblog 的书 —— “WebLog: Publishing Online with Weblogs”。他和 Google 同样相信：“爱”（搜索结果为：75700000）必将战胜“恨”（搜索结果为：7900000）。
- Erik Benson (<http://www.erikbenson.com/>)。
- Justin Blanton (<http://justinblanton.com>) 取得过计算机技术学士学位，现正在硅谷的法学院读书，主要研究知识产权法，以后可能会从事专利起诉和诉讼。他大多数的“空闲时间”都花在了在网站上编写各种东西上，包括 Mac OS X、手机以及其他的小发明、Movable Type CMS 通用技巧手段和日常生活琐事。
- CapeScience.com (<http://www.capescience.com/>) 是一家 Web 服务公司、Cape Clear Software 的开发团体。除了为 Cape Clear 的产品提供支持之外，CapeScience 还生产各种有趣的网络服务产品，包括生活服务等其他服务项目的客户端、工具软件，以及其他一些有意思的玩意儿。
- Antoni Chan (<http://www.alltooflat.com/>) 是 All Too Flat 的创始人之一。All Too Flat 是一个怪诞内容、恶作剧及玩笑幽默的聚集地。从 2001



年10月开始, Antoni Chan 历经一年多时间, 开发出了 Google “镜像”, 总共有 2500 多行 CGI 代码。在网站工作之余, 他喜欢演奏音乐, 打保龄球, 玩完飞盘之后跑跑步。

- Tanya Harvey Ciampi (<http://www.multilingual.ch>) 成长在白金汉郡, 后到苏黎世求学, 在那里她获得了翻译文凭。她目前住在瑞士意大利语区的 Ticino, 从事将意大利语、德语和法语转换到英语的技术翻译及校对工作, 并在她自己的网站 “WWW Search Interfaces for Translators” 上讲授翻译和互联网搜索技巧。业余时间, 她喜欢同父亲一起在爱尔兰西海岸钓鱼, 还喜欢写诗和演奏凯尔特音乐。
- Peter Drayton (<http://www.razorsoft.net/weblog/>) 是微软 CLR 团队的项目负责人。在加入微软之前, 他是 DevelopMentor 的独立顾问和培训师, 著有 “C# Essentials” 和 “C# in a Nutshell” (O'Reilly) 两本书。
- Andrew Flegg (<http://www.bleb.org/>), 毕业于华威大学, 后在英国任职于 IBM 公司。他现在是 Hursley 实验室内部网站的管理员。他目前的主要工作 (也是他的乐趣) 是和 Perl、Java、HTML 及 CSS 打交道。Andrew 特别迷恋简洁的、可重用的代码, 这些代码总能节省大量的开发时间。他参与过几个开源项目, 还为 RISC 操作系统 (类似于 Iyonix PC 上所用的那种 —— Iyonix PC 是第一种应用 Intel Xscaler 技术的桌面电脑) 写过一些商业应用程序。
- Andrew Goodman (<http://www.page-zero.com>) 是 Page Zero Media —— 帮助委托人执行更好的付费搜索项目 —— 的建立者和负责人。他的日常工作是担任 Traffick.com 的总编辑, Traffick.com 是搜索引擎和入口的逆向指导。“Fortune Small Business”, “The Washington Post”, “New Media Age”, “The New York Times”, “Bloomberg Markets”, “Business Week”, “Reuters”, “The National Post”, “CBS Marketwatch”, “Forbs” 和一些其他商业刊物都发表过他关于搜索广告页的观点。他是 “Winning Results with Google AdWords” (McGraw Hill, 2004 年末) 的作者。他最喜欢的 Google hack 是 GooPoetry。
- Kevin Hemenway (<http://www.disobey.com/>), 这位 disobey.com 的创始人更以 Morbus Iff 知名。disobey.com 的建站宗旨是 “为不满意的人提供满意的内容”。他集出版者、开发者、作家于一身, 令人惊讶的是, 他还创作了很多烹饪的书! (流行的开放源代码阅读器 AmphetaDesk、