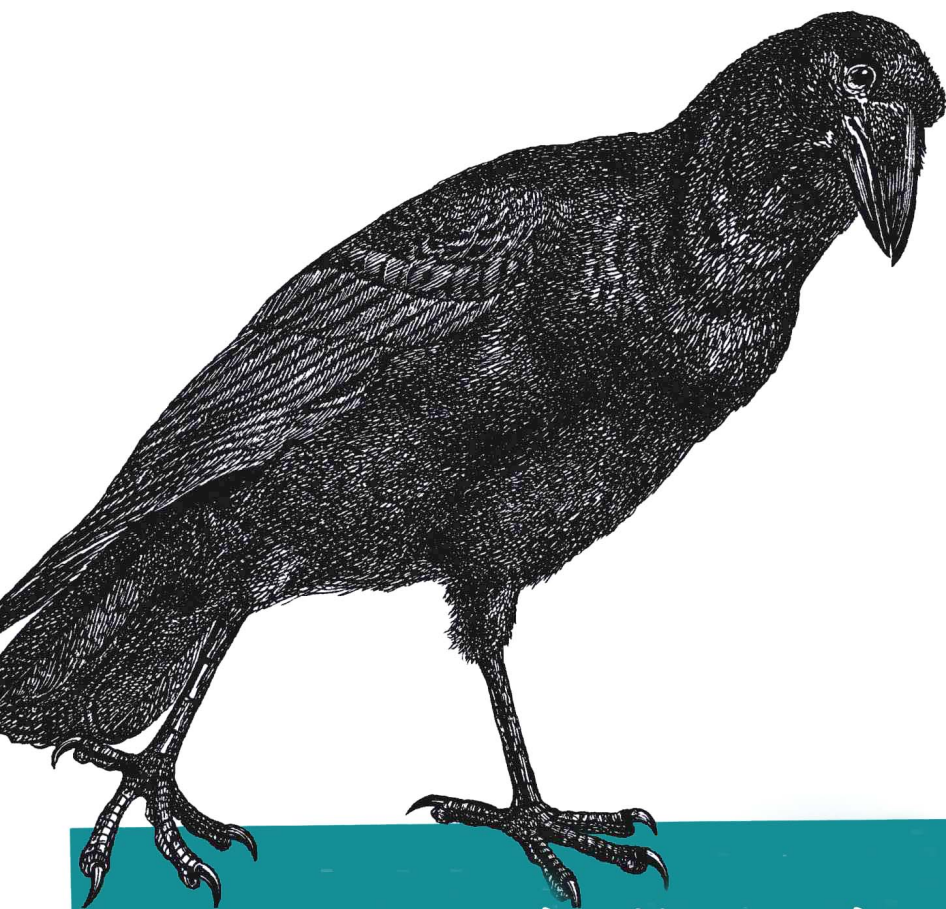


Complete Web Monitoring



网站性能监测与优化

用户行为分析/网站性能分析
/社区管理/竞争对手分析

[美] *Alistair Croll Sean Power* 著

袁菲 房向明 袁继彬 译

O'REILLY®



人民邮电出版社
POSTS & TELECOM PRESS

O'REILLY®

网站性能监测与优化

用户行为分析/网站性能分析
/社区管理/竞争对手分析

[美] Alistair Croll Sean Power 著

人 民 邮 电 出 版 社

北 京

图书在版编目 (C I P) 数据

网站性能监测与优化 / (美) 克鲁尔 (Croll, A.),
(美) 鲍尔 (Power, S.) 著; 袁菲, 房向明, 袁继彬译
— 北京: 人民邮电出版社, 2010.12
ISBN 978-7-115-23497-1

I. ①网… II. ①克… ②鲍… ③袁… ④房… ⑤袁
… III. ①网站—管理 IV. ①TP393.092

中国版本图书馆CIP数据核字(2010)第147498号

版权声明

Copyright©2009 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2010. Authorized translation of the English edition, 2009 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.

本书中文简体版由 O'Reilly Media, Inc. 授权人民邮电出版社出版。未经出版者书面许可, 对本书的任何部分不得以任何方式复制或抄袭。

版权所有, 侵权必究。

网站性能监测与优化

◆ 著 [美] Alistair Croll Sean Power

译 袁菲 房向明 袁继彬

责任编辑 傅道坤

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号

邮编 100061 电子函件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

北京鑫正大印刷有限公司印刷

◆ 开本: 787×1000 1/16

印张: 35.75

字数: 735千字

2010年12月第1版

印数: 1~3500册

2010年12月北京第1次印刷

著作权合同登记号 图字: 01-2009-7052号

ISBN 978-7-115-23497-1

定价: 79.00元

读者服务热线: (010) 67132705 印装质量热线: (010) 67129223

反盗版热线: (010) 67171154

内 容 提 要

本书紧紧围绕业务增长、吸引用户、改善用户体验等目标，全面、翔实地介绍了 Web 监测的方方面面，涵盖了 Web 技术、市场管理、数据分析等内容，将较抽象的业务目标转换成具体的数据、工具以及实践。本书中引用了大量现实的例子和详细的数据，还讲解了许多工具的使用方法。

对于正在维护网站的管理者来说，本书不仅可以帮助他们找出解决方案，还可以提供有益的思路和借鉴。本书也适合所有对 Web 应用感兴趣的人士。

O'Reilly Media, Inc.介绍

为了满足读者对网络和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.授权人民邮电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 20 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，这使其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顶尖级的技术专家。O'Reilly Media, Inc.还有人数可观的固定的作者群——他们本身是相关领域的技术专家、咨询专家，同时编写著作，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，所以它了解市场上真正需要什么图书。

感谢 Kirsten，她有耐心、聪明、善解人意、顽皮，在过去的一年中，她毫无怨言地独自度过了每个夜晚和周末。

——Alistair

感谢爸爸、妈妈和 Mark。你们永远的支持和爱是我能够成功的原因。“谢谢”已经不足以表达我对你们的尊敬、感激和爱。

——Sean

前言

这是一本可以让读者全面了解 Web 可视性的书。现在多数事情都和 Web 有关，如果读者正在从事 Web 相关的工作就需要理解哪些事情会影响 Web。当然，每件事都会有影响，也就是说影响因素已经超出了访问者对网站的访问和相关的竞争者。

虽然 Web 已经成为我们日常生活的一部分，但是 Web 可视性却差强人意。Web 运维人员很少能对访问者与网站交互、Web 资产健康、在线社区如何谈论组织、竞争对手所从事的工作有完整的理解。他们所了解的可视性很零碎、不完整。

Web 运维人员没有必要这样做。最近几年，在监测、搜索、数据收集和报告等方面的进步使得构建一个更为全面的 Web 呈现界面成为可能。

本书将回答下面一些基本问题。

- 访问者在网站上干什么？
- 他们如何采取这些行动？
- 他们为什么开始要这么做？
- 他们这样做是因为宕机还是慢？
- 人们在其他地方如何谈论公司？
- 如何和对手比较？

回答这些问题，需要运用多种手段：Web 分析、用户调查、可用性测试、性能监测、社区管理和竞争性分析。

首先，这看起来令人畏惧：有上百种工具可供选择，从免费的到十分昂贵的；还要追踪上百万的网站和更多的访问者；有许多参与者用他们各自的方式关注你的网站，包括投资者、决策层、销售人员、设计者、开发者和支持团队。

本书希望用各种触手可及的技术来武装读者，帮助读者拟定、实现一个完整的 Web 监测策略。

如何使用本书

本书有 5 个部分。

- 第 1 部分“第 1 章～第 4 章”了解完整 Web 监测的需求。因为每个业务是不同

的，需要查看影响业务的指标和性能。建议通读这一部分，了解 Web 可视性的范围。

- 第 2 部分“第 5 章～第 7 章”关注用户的网站行为。不仅是简单的分析，还包括诸如可用性、记录交互过程和对访问者需求的调查。每章都提供了技术介绍和历史、思想框架、实现细节，并介绍如何在组织内部分享数据。当读者需要某种可视性时，可以再来翻阅这些章节。
- 第 3 部分“第 8 章～第 10 章”关注 Web 性能监测。这是前两部分的补充，看访问者可能做些什么事情，而不是他们做了什么。了解 Web 性能基础后，本书会介绍综合测试和真实用户监测 (RUM)。想了解 Web 性能，可以阅读第 8 章；而想了解关于性能监测，可以阅读第 9 章和第 10 章。
- 第 4 部分“第 11 章～第 16 章”关注 Internet 的其余部分，包括社区 (内部和外部)、竞争者。这部分将介绍在线社区的出现以及公司如何利用它们。这部分提供了分析社区的框架，介绍了 8 种通用社区模型和监测方法。然后，这部分将快速介绍社区内部。最后，这部分将介绍如何用监测自己网站的工具来追踪竞争者。如果读者从事社区监测、竞争力分析方面的工作，建议通读这部分。
- 第 5 部分“第 17 章～第 18 章”讨论了如何整合和加强多种监测数据，以获得完整视图。然后展望 Web 监测的未来，并邀请读者一起在线讨论。

本书包括什么，不包括什么

本书包括的内容十分广泛。本书试图提供关于每种监测方法的基础的、摘要性的内容，以及具体的例子和实际执行建议。

就某些主题，有许多书比本书挖掘的要深入。如果是这样，那么本书将会明确指出。

本书不是关于网站建设的书，也不会让网站变得更快、更易用，也不会让人们更多地谈论你。但是，本书将介绍如何判断网站是否快速、有效、可用性好，并了解人们是否在 Web 上谈论你。

适用人士

如果读者负责报告和改进在线业务，那么这本书就适合你。听起来好像适合很多人，实际上就是这样：全面的 Web 可视性涉及多个学科。但是，这并不意味着学科应该相互脱节，可以考虑下面的情况。

- 你负责 Web 分析，需要了解性能、社区推荐以及可用性对访问者的影响。
- 你负责管理 Web 基础设施，需要知道如何监测流量突然的增加、如何建立模拟真实用户访问的测试。
- 你正管理社区，需要知道什么样的行为会增加网站的业务量。

- 你负责竞争性分析，需要了解网站与竞争对手的比较情况。

企业家和营销管理者可以用本书中的信息定义组织成功的标准，本书还适用于那些对此标准负有责任的产品经理、Web 运维人员、开发者、系统管理员。

必须了解

本书假设读者了解基本的 Web 元素——URL、搜索引擎、Web 导航，以及社会化网络。你不需要了解 Web 应用内部细节，但是你要理解它们可能会以不可预测的方式宕掉。你知道网站可以更快，但是你不能准确地了解它有多慢或者什么是用户关心的。

最后，你需要知道 Web 驱动业务。

使用代码样例

这本书有助于完成工作。一般，读者可以在程序和文档中使用本书的代码。除非读者复制了代码的重要部分，否则读者不需要联系我们以获得许可。例如，在程序中使用几段本书中的代码，不需要获得许可。销售和分发配套例子光盘需要获得许可。引用本书和代码样例回答问题不需要获得许可。把本书中的代码样例整合到产品文档中需要获得许可。

本书赞成注明来源，但不是必需的。来源通常包括标题、作者、出版者和 ISBN。

如果读者有版权方面的问题，可以联系 permissions@oreilly.com。

如何联系我们

Web 监测变化非常快。正如读者将在本书中看到的，许多因素（包括移动性、客户端应用、新的浏览区、视频和社会化网络平台的风靡）使维护良好的 Web 可视性变得困难。

虽然本书只是完整 Web 监测的基础，但是本书的合作网站 www.watchingwebsites.com 提供了类似内容、指向工具的链接、论坛等，还可以通过 authors@watchingwebsites.com 联系我们。

厂商策略

本书中有大量的例子和截图。许多厂商提供给我们访问其系统的机会，我们免费使用了这些工具。但是，不要把例子认为是广告。在许多情况下，本书使用已经获得的免费工具。也有某些公司有许多功能不同、用途广泛的工具，在本书中显得比较突出，特别是 Google。其他公司有替代方案。

维护 www.watchingwebsites.com 的一个原因是为了接触 Web 监测行业成熟和壮大过程中的新人和了解不可避免的错误。

审阅者

这本书是一群人努力的结果。Web 监测是一个复杂的话题，本书采纳了许多行业专家的建议。本书特别希望能够从厂商得到反馈，因为本书的作者都为厂商工作，而本书期望内容是中立和客观的。

错误和疏漏是作者的错，聪明的想法和妙语归功于审阅者。要是没有下面的工作，可能也没有本书。

- Imad Mouline 用完全独立于厂商的观点审阅了本书的第 1 部分。
- Bob Page 使我们不止一次脱离困境，并提供了极佳的反馈。
- Hooman Beheshti 有渊博的 Web 性能和嗅探器知识。
- Matt Langie 和 Sean Hammons 为 Web 分析章节提供了很好的反馈。
- Bryan Eisenberg 提供了历史教程，并为 A/B 测试、实验提供了很好的反馈。
- Avinash Kaushik 总是能够问到点子上，并能做好充分准备。
- Stephanie Troeth 能够从结构上把握整体，是他说服我们重新构思和编写了 Web 交互分析和顾客心声章节。
- Clicktale 的 Tal Schwartz 和 Kampyle 的 Ariel Finklestein 提供了很好的素材和反馈，并继续创新着。
- iPerceptions 的 Johnathan Levitt 提供了对顾客心声一章的深刻见解。
- Foresee 团队，特别是 Eric Feinberg 为我们介绍了 Lee Pavach 和 Larry Freed。
- Tealeaf 的 Robert Wenig 和 Geoff Galat 对本书的公正性做了确认，并提供了其博客上的丰富资源。
- Coradiant 的 Fred Dumoulin 审阅了 Web 性能和真实用户监测部分，并继续进行着 Web 监测方面的工作。
- Doug McClure 为本书提供了 ITIL、BSM 和 IT 监测方面的一般性经验。
- Symphoniq 的 Joe Hsy 为真实用户监测章节提供建议，并丰富了收集方法学。
- Akamai 的 Tim Knudsen 提供了真实用户监测和内容分发网络方面的反馈。
- Oracle/Moniforce 的 Daniel Schrijver 提供了非常好的最终反馈。
- 感谢 Gomez 的团队，特别是 Imad Mouline 和 Samantha McGarry。
- 感谢 Keynote 的团队，特别是 Vik Chaudhary 和 David Karow。
- 感谢 Webmetrics 的团队，特别是 Peter Kirwan、Arthur Meadows 和 Lenny Rachitsky。

- 感谢 AlertSite 的团队，特别是 Richard Merrill 和 Ken Godskind。
- Radian6 的 David Alston 审阅了整个社区部分章节，提供了很好的反馈。
- Scout 实验室团队带来了公正、洞察力和幽默。
- Best Buy 的 Steve Bendt 和 Gary Koellig 提供了很好的反馈。
- Stephen Pierzchala 有出色的评论和观察力。
- 来自 Compete.com 的 Lauren Moores 为竞争性监测章节提供了很好的素材，并提供了 Compete 的工具和数据。
- Beth Kanter 用一个下午和晚上的时间审阅了社区部分，并提供了出色、充实的评论。对未能列出的人，本书也一并表示感谢。

致谢

没有耐心和许多人的投入，就没有这本书。我们特别感谢以下人员。

- Networkshop、Coradiant、Bitcurrent、GigaOm、Techweb、FarmsReach、Syntenic 和 Akoha（特别是 Austin Hill）忍受了我们的打搅，并为我们提供了所需的监测数据。
- Interop、Web2Expo、Mesh、Enterprise 2.0、eMetrics、Podcamp、Democamp 和 Structure 为本书提供了技术会议、博客、行业动态等方面的原创资料。
- 为我们提供讨论场所的饭店、酒店、咖啡店，特别要谢谢 Lili & Oli、The Burgundy Lion、Brutopia、the Wired Monk、JavaU、Toi、Moi et Caf 和 Laika 的职员。
- 所有为 Web 开发提供免费工具的人。现在有数以百计的插件、网站、客户端、搜索引擎和报表平台，在本书中我们使用了许多。没有这些工具——公开提供下载，免费服务——Web 将更慢、可用性更差、更不可信赖。我们感谢这些开发者和公司。
- 许多行业巨头提供了他们的想法，为本书铺平了道路，包括 Avinash Kaushik、Paul Mockapetris、Lenny Heymann、Steve Souders、Guy Kawasaki、Ross Mayfield、Hooman Beheshti、Imad Mouline、Bob Page、Cal Henderson、Jim Sterne、Bryan Eisenberg、Eric Reis、Dave McClure、Eric T. Peterson、Charlene Li 和 Jeremiah Owyang。
- 使我们能够保持清醒的艺术家和 DJ，特别是 Underworld、Simon Harrison、Pezznor、Dirk Rumpff 和 Freerange 的其他人。你们说着我们的语言，并使我们的漫漫长夜的写作中保持清醒。
- O'Reilly 的团队，包括 Simon、Lauren、Amy、Marlowe、Sumita、Rob、Rachel 和 Mike，还有 Tim。谢谢他们的信任和耐心。

Alistair 谢谢 Sean、Ian、Hooman、Dan、Marc、Oliver 和 Interop 的其他工作人员；Bitnorth 的每个人；Jesse、Brady、Jen 和 O'Reilly 的其他工作人员；Phils、Cheryl、Paulette、

Michelle、Gordon 和 Kempton Resort 的其他工作人员；Brett、Paul 和 Randy；当然还有 Doreen、Becky 和 Adam。

Sean 谢谢 Alistair、A-K、Brig、Deww、Ian、Hazel、Jayme、Kirsten、Marc、Pablo、Scott、Simon 和 Wes 的出现和无私。

感谢 Cary Goldwax、Dan Koffler、Steve Hunt、Scott Shinn、Guy Pellerin、Tara Hunt、James Harbottle、Lily、Oliver & Angela、Brad & Am、Chris & Melissa、Claudine、Steve Fink、CR、the Houghs and Foucault、Phil Dean、Dennis、Naomi、the shade in between、Adam & the choopa 的团队和家人。

最后，我们真正想感谢的人群是读者。我们的 Twitter 跟随者、Facebook 好友，虽然我们现在是陌生人，但很希望认识你们。欢迎联系我们（Twitter 上的@seanpower 和@acroll）。正是有了这些联系，才有了本书和知识的分享！谢谢。

目录

第 1 部分 Web 监控技术

第 1 章 为什么要观察监测 Web 网站	2
1.1 当前 Web 监测现状	3
1.2 旧的不去，新的不来	3
1.3 隐私问题：追踪他人	4
第 2 章 网站处在怎样的行业中	6
2.1 媒体网站	7
2.2 交易网站	8
2.3 协作网站	9
2.4 软件服务应用网站	11
第 3 章 观察什么指标	13
3.1 用户是否为网站带来收益	13
3.1.1 转化和放弃	14
3.1.2 广告点击	15
3.1.3 离线活动	15
3.1.4 用户发布内容	17
3.1.5 用户订购	18
3.1.6 账号使用	18
3.2 网站流量从何而来	18
3.2.1 链接源网站	19
3.2.2 来自社会化网络的内部链接	19
3.2.3 访问动机	21
3.3 如何做得最好（最坏）	21
3.3.1 网站效率	21

3.3.2	广告营销效率	22
3.3.3	搜索和查找效率	22
3.3.4	问题处理和升级	23
3.3.5	内容流行度	23
3.3.6	可用性	23
3.3.7	用户使用效率	24
3.3.8	社区评价和回报	24
3.4	网站与用户关系如何	24
3.4.1	忠诚度	25
3.4.2	注册	25
3.4.3	访问网站	25
3.5	网站基础设施健康状态	26
3.5.1	可用性和性能	27
3.5.2	服务水平协议承诺	27
3.5.3	内容分发	28
3.5.4	突发流量和网站处理能力	30
3.5.5	性能对业绩的影响	30
3.5.6	营销带来的流量	32
3.5.7	季节性特征	32
3.6	如何与对手竞争	33
3.6.1	网站知名度和排名	33
3.6.2	用户如何发现竞争对手网站	35
3.6.3	相对网站性能	36
3.6.4	对手网站活动	37
3.7	有哪些风险	37
3.7.1	捣乱和垃圾信息	37
3.7.2	版权和法律责任	39
3.7.3	欺诈、隐私和账户共享	39
3.8	用户如何评价网站	40
3.8.1	网站名声	41
3.8.2	趋势	42
3.8.3	社会网络活动	42
3.9	网站和内容在其他地方如何被使用	43
3.9.1	API 访问和使用	43
3.9.2	混合应用、盗版和非法使用	44

3.9.3 与旧系统的整合	45
3.10 随手可及的工具	45
3.10.1 收集工具	45
3.10.2 搜索系统	45
3.10.3 测试服务	46
第 4 章 4 个主要问题	47
4.1 他们做了什么	47
4.2 他们是如何做的	48
4.3 他们为什么这样做	49
4.4 他们能做到这一点吗	49
4.5 把所有的东西结合起来	50
4.6 正确地分析数据	51
4.6.1 始终比较	51
4.6.2 对所有数据进行分类	51
4.6.3 不要满足于平均值	52
4.7 完全 Web 网络监测成熟度模型	53
4.7.1 1 级: 技术细节	53
4.7.2 2 级: 关注自己的网站	54
4.7.3 3 级: 关注 Internet	54
4.7.4 4 级: 建立关系	55
4.7.5 5 级: Web 业务战略	55
4.7.6 监测 Web 站点成熟度模型	56
 第 2 部分 Web 分析、可用性、客户心声	
第 5 章 Web 分析	60
5.1 处理知名度和距离	61
5.2 Web 可视性的核心	61
5.3 分析历史一瞥	62
5.3.1 从 IT 到市场	64
5.3.2 从点击到页面: 追踪到达路径	65
5.3.3 从页面到访问: cookie 的出现	66
5.3.4 从访问到结果: 追踪目标	70

5.3.5	从技术到含义: 标记内容	73
5.3.6	综合视图	75
5.3.7	场所和任务	78
5.4	分析的 3 个阶段	80
5.4.1	寻找网站: 长形漏斗	80
5.4.2	使用网站: 追踪访问者	89
5.4.3	离开网站: 喜忧参半	95
5.4.4	期望的结果	97
5.5	实施 Web 分析	100
5.5.1	定义网站的目标	100
5.5.2	设置数据采集	102
5.5.3	设置过滤	117
5.5.4	确定需要分析的类别	118
5.5.5	标记相关网页内容	118
5.5.6	营销整合	122
5.5.7	测试和验证	124
5.6	共享分析数据	127
5.6.1	重复一致性	128
5.6.2	开始测试	128
5.7	选择分析平台	132
5.7.1	免费平台和付费平台	132
5.7.2	实时分析和趋势分析	132
5.7.3	托管分析和内部分析	133
5.7.4	数据可移植性	133
5.8	前期工作	133
5.8.1	你可以轻松得到什么	133
5.8.2	做一点工作你会得到什么	134
5.8.3	再多做一点工作你会得到什么	134
5.8.4	大量的工作会为你带来什么	134
5.9	Web 分析成熟度模型	134
第 6 章	监测 Web 可用性	136
6.1	Web 网页设计就是假设	136
6.2	浏览内容: 滚动行为	138
6.3	正确交互: 点击热点图	142

6.3.1	可用性和功能可见性	142
6.3.2	分析鼠标互动	144
6.4	数据输入和丢弃：表单分析	147
6.5	单个访问：重放	149
6.5.1	高效捕捉：重放的内容取决于要解决的问题	149
6.5.2	回溯分类：回答“如果……会怎样”	157
6.6	实施 Web 交互分析	159
6.6.1	了解设计假定	159
6.6.2	确定收集内容	160
6.6.3	部署和收集交互信息	160
6.7	一些相关问题	165
6.7.1	如果网页发生变动会怎样	165
6.7.2	Web 交互分析无法看到的用户行为	166
6.7.3	动态命名和页面上下文	166
6.7.4	浏览器渲染问题	167
6.7.5	不同的点击次数有不同的含义	167
6.7.6	Web 交互分析对性能的影响	168
6.7.7	回放、页面中性化和插件	168
6.7.8	隐私	169
6.8	Web 交互分析成熟度模型	171
第 7 章	客户心声	172
7.1	旅游业的困境	172
7.2	用户没有按照网站意图访问	174
7.3	什么是客户心声	175
7.3.1	透彻了解和线索	176
7.3.2	主观评分	177
7.3.3	人口统计	177
7.3.4	环境统计	178
7.3.5	监测不能跟踪来源的数据	179
7.4	什么是客户心声不能做的	180
7.4.1	它不能替代其他形式的收集	181
7.4.2	它不能代表用户基础	181
7.4.3	它不能代替社区	181