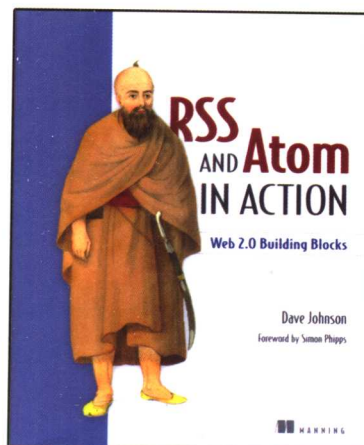




RSS AND Atom IN ACTION: Web 2.0 Building Blocks

RSS AND Atom IN ACTION 中文版

Foreword by Simon Phipps



[美] Dave Johnson
李秀忠 译 著



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

RSS and Atom in Action 中文版

Rss and Atom in Action

[美] Dava Johnson 著

李秀忠 译

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书以虚拟的 FinModler 技术公司在他们的财务建模软件中, 如何使用博客和播客技术为主线, 贯穿了 newsfeed 格式、newsfeed 解析, newsfeed 的提供、Atom 发布协议, 以及 10 个活学活用的博客小应用程序等内容。本书的最后又把这些精彩内容进一步延伸, 向你介绍了一些出人意料的想法, 这些想法足以起到一个抛砖引玉的作用, 为你更深入的使用这些技术带来启发。适合面向使用博客和播客技术进行开发的程序员, 希望了解 WEB2.0 最新技术发展方向的项目管理人员, 想通过博客, wiki、newsfeed 技术推广网站的网站管理员。

1-932394-494 RSS and Atom in Action for Programmer Testing by DAVE JOHNSON Original English language edition published by Manning Publication Co., 209 Bruce Park, Avenue, Greenwich, Connecticut 06830. Copyright © 2006 by Manning Publications Co.. Simplified Chinese Language edition copyright © 2007 by PUBLISHING HOUSE OF ELECTRONICS INDUSTRY. All Rights reserved.

本书中文简体版专有出版权由 Manning Publication Co. 授予电子工业出版社。未经许可, 不得以任何方式复制或抄袭本书的任何部分。

版权贸易合同登记号 图字: 01-2007-3444

图书在版编目 (CIP) 数据

RSS and Atom in Action 中文版 / (美) 约翰逊 (Johnson, D.) 著; 李秀忠译. —北京: 电子工业出版社, 2007.11

ISBN 978-7-121-05116-6

I. R… II. ①约…②李… III. ①主页制作②网站—开发 IV. TP393.092

中国版本图书馆 CIP 数据核字 (2007) 第 150734 号

责任编辑: 梁 晶

印 刷: 北京市天竺颖华印刷厂

装 订: 三河市金马印装有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

经 销: 各地新华书店

开 本: 787×980 1/16 印张: 24.75 字数: 510 千字

印 次: 2007 年 11 月第 1 次印刷

定 价: 49.80 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

译序

博客和播客技术确实已经占据了互联网的各个角落。当你读到这句话的时候，可能你身边的人正在构思一篇博客文章，捕捉自己的灵感，沉淀自己的思想。博客和播客技术正在日益走向成熟，它更加利于 Web 创作和分享，既在张扬着每个个体的个性，又在展现着群体的力量。这一切都要感谢那些博客技术的先驱者。

对博客技术的变迁，我的感触也很深刻。自己曾在博客刚刚出现时，就成为博客中的一员，感触最深的就是曾在博客上上演过“蚂蚁搬家”的故事。

第一个博客可以说是对博客技术的初体验。兴致勃勃地去对比各大博客网站所提供的吸引眼球的服务，最后确定在一个博客网站上安家落户，但在使用过程中，每每发布一篇博客都要被网站强制加上版权声明和插播一条广告，人就觉得很很不愉快，一个宁静的心情就这样被打扰了，于是开始了第一次举家搬迁，幸好自己的博客文章还不是很多。于是又选择了一个博客网站安营扎寨，而且写了很多文章，但由于还留恋旧地，于是同时维护着两份博客，就像书中描述的 Rangu 那样。

可是，好景不长，也许是博客技术发展太快了，这个博客网站选择了技术升级，结果在免责声明中，当我别无选择地按下了“接受升级”的按钮之后，就注定了我永远进不了“家门”了，我的博客升级失败，后台的管理页面进不去了。气愤之下，我被迫选择了第二次搬家。这次选择博客网站，我更看重博客网站的备份和导入导出功能。万一哪天博客网站关闭了，我不至于流离

失所，家财尽失，至少我可以很快地重建家园。这次迁移是比较痛苦的，需要重建分类，把以前的文章一篇一篇从旧网站上拷贝下来，发布到新博客中去，像一只辛苦的蚂蚁。看着自己在新博客网站上当月博客发布排名的飙升，怎么也高兴不起来。自己几年的生活感触，就被定格到一个月中，早已没有了博客的味道。

这些惨痛的经历就是由于博客技术的不成熟造成的。没有统一的规范，没有统一的数据交换格式，没有统一的编程接口，没有……

然而，当我看到这本书的时候，心情为之一振，我所期待的这一切正悄悄地走来，博客界正在发生着翻天覆地的变化。如果你的网站还在固守着旧的思想和技术，可能很快就会被淹没在这场变革中。书中讲述的这些技术，就是我们每一位博客，甚至是播客，以及每一位致力于博客技术的开发者所需要的东西，而且作者的思想，也将点燃你创作的灵感。

作为译者，非常渴望在您和作者之间架起一道桥梁，让您和作者进行心与心的沟通，把作者的思想直接传递给您，而感觉不到译者的存在。但是由于个人能力所限，还不能达到这种境界，而且纰漏之处，也在所难免，非常真心地希望得到您的批评和指正。

李秀忠

2007-7-8 于深圳

E-mail: xiuzhong.li@gmail.com

前言

preface

无论你认为第一份博客是始于大约 1997 年 Jorn Barger 命名的“weblog”在线杂志，还是 Web 诞生后不久由 NCSA¹和 Netscape 创建的“what’s new”的网页，或者是美国独立战争时期的政治传单。你不得不承认博客已经不是一个全新的概念。博客仅是代表在线写作的一个词汇而已。

那么什么是最新的呢，那就是 newsfeed 和发布协议（publishing protocols）等博客技术在 Web 上的广泛应用。在 20 世纪 90 年代末，博客软件以及 Web 门户的开发者需要一种标准的数据格式，以便他们能够轻松地联合 Web 上的内容。因此，RSS、Atom 以及其他 XML 的 newsfeed 格式就诞生了。同时，他们还需要一套可以针对 Web 进行编程和发布的标准协议。因此，XML-RPC、SOAP 和 Web 服务也应运而生。

如今，真的要感激人们对博客、podcasting、wiki 表现出来的热情，而且那些方便开发的博客技术也已经无处不在。Newsfeed 不仅是博客的标准特性，而且也是网站、搜索引擎乃至无处不在的 wiki 的标准特性。电脑、音乐播放器，甚至移动设备都捆绑了 newsfeed。Newsfeed 技术已经成为浏览器、办公软件、操作系统的一个标准组成部分。即使你没有看到这种革新的机会，你的客户都可能要求你提供这些技术，现在是该做准备的时候了。

¹ 译注：NCSA，国家超级计算机应用中心（National Center for Supercomputer Applications）。

本书向你介绍了使用这些博客技术建立应用程序的知识。考虑到读者当中，会有些开发者对我们前面的讲述不以为然，所以，我们会先从一些博客应用的故事开始，向你展现一个应用博客技术的全新的协作方式。之后，我们会向你解释你需要了解哪些博客技术——不光是 RSS 和 Atom。我们也会向你讲述博客服务器结构、博客 API 和 Web 服务协议。

为了帮助你从头开始，我们涵盖了博客技术开发工具包，完整的博客服务器，newsfeed 解析器，博客客户端库，以及第 2 部分 10 个可以立即使用的用 Java 和 C# 写的博客应用程序，或者说是**博客小应用程序**（blog apps）。我们的博客服务器和这 10 个应用程序，即 *Blogapps server* 和 *Blogapps examples*，是作为一个开源项目在 <http://blogapps.dev.java.net> 上进行维护的，当然也欢迎你帮助我们一起维护和改进这个项目。

我希望我们已经为你提供了开始建立优秀的博客应用程序所需要的一切，更期待着能够看到你建立的一些应用程序。最后，希望你能够喜欢这本书！

序

foreword

自从 Henry Ford¹告诉他的消费者，他们可以拥有“任何颜色的（T 型车），只要它是黑色的，”以来，我们的消费者群体就被极少数创造者的目标和视野驱动着。但是到了二十世纪九十年代，万维网的出现导致了 Internet 的爆炸性普及，而且凌驾于消费者群体之上，观念的单方灌输也已成为过去的回忆。无处不在的 Internet 立即将我们无情地推进了一个崭新的时代，在这个时代中，观念的单方灌输已经变成了多方交流，而且更关键的是，消费者群体也由被动的消费变成了积极的参与。

就在这样的背景下，我和一些同事为 Sun Microsystems 启动了 blogs.sun.com 这个网站。我们可以预见在这个“参与的时代”，公司成功的关键将是为 Sun 的职员提供一种方式让他们直接面对技术和他们所参与的客户社区。全世界，不管你生活在哪个角落，大家都会达成这样的共识：今天的博客将会像二十世纪九十年代初的网站一样迅猛成长。

伴随着博客，几乎是同时带来了另外一种可能对社会产生更大影响的技术：联合 feed（syndication feed），一个计算机可读的博客内容列表。如今它被用于博客阅读程序以及聚合器

¹ 译注：Henry Ford 亨利·福特，美国汽车大王，1903 年，创立福特汽车公司。福特汽车公司生产的大多数 T 型车都是黑色的，因此，亨利·福特曾有句名言“任何顾客都可以把他的汽车喷涂成他想要的颜色，只要汽车是黑色的”（Any customer can have a car painted any color that he wants so long as it is black）。实际上，福特汽车公司也曾生产过其他颜色的汽车，但后来，为了提高生产效率和降低涂料成本，福特汽车公司只使用价格低廉且易于干燥的黑涂料来批量生产黑色汽车。

（像 BlogLines²网站或者我用来建立摘要博客“The Daily Mink”³的 Planet Roller⁴聚合器）中。联合 feed 允许对博客内容进行创新性的再利用，并且为内容共享开启了一片新天地，例如播客技术。尽管对联合 feed 的使用还处在初级阶段，但是，我可以预言这样的一个重大事情，那就是，随着建立和使用它们的能力的提高，它必将被嵌入到我们使用的计算机和移动设备的操作系统中。

联合 feed 可能看上去很简单，但是无论是 RSS 格式还是 Atom 格式的联合 feed，它们都是 Web 和 XML 核心体系进化过程中的重要阶段。XML 的原创者把 XML 看作是允许使用树状结构表现文档的通用文档语言。联合 feed 却把另外一个更强大的结构带到了 XML 中来，那就是——列表和集合。

列表和集合（例如数据库）已经处于许多计算的核心，而且联合 feed 为编程提供了一种有组织地共享数据的手段。它们为轻松实现 SOA（service-oriented architectures，面向服务的体系结构）提供了一种途径，并且开启了对围绕在我们身边的所有数据（像银行帐号、健康记录、账单信息、旅游历史，等等）的富有想像力的应用之门。联合 feed 使我们的 Web 是可编程的，而且不仅仅是这些，Atom 规范化了访问 feed 的手段，并且提供了一个 API 来降低网站和使用它的 feed 的程序之间的耦合。

在“Web 2.0”运动中，蜂拥而至的人们开始使用联合 feed 以及 Ajax 来建立网站，诸如 Flickr，del.icio.us、Bloglines 以及 Technorati 等等，而他们也仅是这些应用当中的冰山一角。

本书对那些想要为未来做准备的人来说是一本重要的参考书。为了获得博客技术方面的信息，你可能已经拾起了它，但是它所提供的远非如此。它是未来的发射台，有了它，你就不必像 Tim Bray、Sam Ruby、Dave Winer 以及 Mark Pilgrim 这些开拓者那样，不得不在前行的过程中面对一切。

这本书就是为你准备的，它教给你的这些技能，是开启将要改变这个世界的“参与时代”编程的钥匙。读下去吧，它将使你更聪明地编程，并开启未来！

SIMON PHIPPS

首席开放源码官（Chief Open Source Officer）

Sun Microsystems, Inc.

² <http://www.bloglines.com>.

³ <http://www.webmink.net>.

⁴ <http://www.rollerweblogger.org>.

致谢

acknowledgments

尽管封面上只有我一个人的名字，但是，在幕后有很多人为这本书的问世作出了努力，我应该为此真诚地感谢他们。

先从 Rick Ross 开始吧，是他鼓励我写作，并且把我介绍给 Manning 出版社和出版者 Marjan Bace。和 Manning 出版社的合作一直是非常愉快的。在这里感谢 Denis Dalinnik、Jody Gilbert、Mike Levin、Dottie Marsico、Sharon Mullins、Frank Blackwell、Mary Piergies、Karen Tegtmeyer、Helen Trimes，以及其他全体人员。

同样感谢审阅者 Tim Bray、Simon Brown、Steven Citron-Pousty、Rick Evans、Jack Herrington、Frank Jania、Lance Lavandowska、Robert McGovern、John Mitchell、Jaap van der Molen、Yoav Shapira、Doug Warren、Henri Yandell、Peter George、Paul Kedrosky、Joe Rainsberger、Pim Van Heuven、Patrick Chanezon、Alejandro Abdelnur，以及 Walter Von Koch，是他们在最初审阅这本书的时候提出了很多宝贵的建议。并且特别感谢 Mike Levin，是他为我的最终手稿进行技术校对。

感谢 Simon Phipps，是他为本书作序，并且在他的私人网站上勇敢地使用了本书中的软件。同样感谢 Masood Mortazavi，是他提供了第 1 章第一个屏幕截图中关于“风险价值”方面的文本。

关于本书

about this book

本书向开发者介绍了如何使用博客技术建立应用程序。第1部分介绍了博客技术的基础知识,包括博客和 wiki 服务器的结构,RSS 和 Atom newsfeed 的格式,MetaWeblog API 以及 Atom 协议。一旦具备了这些基础之后,我们会把焦点放在建立应用程序上面。在第2部分,每一章都讲述了一个可以立即投入使用的博客应用程序。

你会在第0章“你要先具备哪些知识”中,找到本书更详细的路线图和介绍。

预期读者

本书的预期读者是开发者和需要了解博客、wiki 以及 newsfeed 的 IT 技术革新者。如果你想在你的程序中加入 newsfeed 的读取能力或者在你的网站上加入 newsfeed 的生成能力,那么这本书就是为你准备的。如果你想完成面向 Web 的自动化发布过程,你会发现本书对你很有用。如果你要部署博客或 wiki,在选择软件之前需要了解博客和 wiki 服务器的结构,那么,你会在这里找到答案。如果你仅仅想找一些新想法、新机会,你也会在这里发现一些有价值的东西。

对于本书的大部分章节,我们都假设你对使用 Java 或 C#进行 Web 开发有所了解。要了解更多的关于本书需要的预备知识和本书内容的完整路线图,请阅读第0章,那里会向你介绍要先具备哪些知识。

相关下载

本书的所有源代码都可以在 Java.NET 上在线下载，并且作为一个开源项目 Blogapps 进行维护。每章的例子都是独立打包的。你可以使用 Ant 构建 Java 例子，但是对于 C# 和 ASP.NET 的例子则需要 Microsoft Visual Studio。你可以在 Blogapps 项目网站 <http://blogapps.dev.java.net> 上找到构建和运行每个例子的完整指南。

Blogapps 服务器是一个完整的博客和 wiki 服务器，支持本书中讨论的所有 newsfeed 格式和发布协议。在第 2 章我们会介绍如何下载、安装以及启动 Blogapps 服务器，在上面提到的网站上同样可以下载到 Blogapps 服务器。

你也可以从出版商的网站 (www.manning.com/dmjohnson) 上得到本书的源代码。

印刷约定

对于 Java、C#、XML 的源代码清单中的类名、常量以及其他单词，我们都使用 Courier 字体。在源代码清单中需要高亮显示的重要片段我们都使用**粗体 Courier**。我们使用像●这样的“球形提示符”来指出源代码清单中我们要特别讲解的代码行。

作者在线

只要购买了 RSS and Atom in Action 一书，你就同时可以自由访问 Manning 出版社的专用 Web 论坛，在那里你可以发表对本书的评论，提出一些技术问题，并可以从作者以及其他读者处获得帮助。要进入论坛或订阅相关信息可以访问 www.manning.com/dmjohnson。一旦注册，这个网页会提供更多的信息，包括如何进入这个论坛，可以获得哪些帮助以及论坛的规则。

Manning 对读者的承诺是提供一个可以在读者之间以及读者和作者之间进行有效沟通的场所，但并不承诺作者的参与程度，因为作者对论坛的参与是自愿的，而且是不计报酬的。我们建议你向他们提出一些挑战性的问题，免得他们失去热情！

只要本书投入印刷,“作者在线”论坛以及前面讨论的文章就可以在出版社的网站上访问了。

关于作者

DAVE JOHNSON 就职于 Sun Microsystems, 负责开发, 支持和推广博客技术。在加盟 Sun 公司之前, Dave 曾在多家软件公司工作过, 包括 SAS 协会、HAHT 商业公司、Rogue Wave 软件公司。在 2002 年, 不满足现状的他想在工作中建立更酷的软件, Dave 每个夜晚和每个周末都在工作, 终于开发出了一个名叫 Roller 的开源博客服务器, 至今在 Sun、IBM 和 JRoller.com 上有成千上万的博客在使用。

关于书名

In Action 丛书结合了引导、概述, 以及教你如何做的实例, 帮助你学习并记忆。根据对认知科学的研究, 人们记忆的事情, 往往是那些他们在自我激励的探索中发现的事情。

尽管 Manning 没有人是认知科学的科学家, 但我们确信学习是一个持续的过程, 它必须经历探索、行动、感兴趣, 到重复已学知识的各个阶段。人们理解和记忆新事物时, 只有在积极地探索之后, 才能称得上是精通了它们, 人们是在实践中学习的。In Action 指导原则的本质就是实例驱动。它鼓励读者发掘新事物、创作新代码、探索新想法。

关于本书的书名还有一个更现实的原因就是: 我们的读者都很忙。他们因工作需要或者为了解决实际问题而寻求书籍的帮助。他们需要的是可以让他们快速上手的书, 他们需要的是在想学的时候就能学到想要的东西的书, 他们需要的是能够帮助他们实践的书。本丛书就是为这样的读者准备的。

关于封面

RSS and Atom in Action 的封面是一个“叙利亚的修道士”。穆斯林修道士生活在宗教社区, 和基督修道士一样, 退居世外, 过着清贫和潜心修教的平静生活。他们被认为是智慧、医术、诗歌、启蒙、格言的源泉。这幅插图摘自《土耳其帝国时期的服装收藏》, 这幅画册由住在伦敦 Old Bond 大街的 William Miller 出版于 1802 年 1 月 1 日。由于画册的扉页已经缺失, 我们无法考证它的具体日期。这本画册的目录使用英语和法语两种语言来说明这些插图, 而且每幅插图都有两位画家的名字。毫无疑问, 他们一定会很惊奇, 两百年后, 他们的作品居然荣幸地出现在计算机编程书籍的封面上。

这幅画册是 Manning 出版社的编辑在一个位于曼哈顿西 26 街车库里的古玩跳蚤市场上买到的。卖主是一个居住在土耳其安卡拉的美国人，进行交易的那天他正在收摊。Manning 出版社的编辑身上没有带足够的现金，想要使用信用卡和支票付账都被卖主婉言拒绝了。如果晚上卖主飞回安卡拉，交易将会化为泡影。那后来怎么办呢？他们找到的最有效的解决方式，居然是最原始的握手成交的口头协议。卖主率直地提出可以把钱寄给他。这位编辑竟然带着写着卖主账户信息的纸条，腋下夹着那幅画册走出了跳蚤市场。不用说，我们当然是在第二天把钱转了过去。时至今日，我们仍记忆犹新，仍然对一个陌生人对我们的信任心存感动。

来自土耳其收藏画册的这幅图，和其他插图一样出现在我们的封面上，把两个世纪以前各种各样丰富多彩的民俗服饰带到我们的生活中来。它唤起了我们对那个历史时期（以及除了沟通极为方便的现代之外的其他各个历史时期）的隔离感和距离感。

从那以后，服饰风格已经改变了，当时富于地域特色的多样性，都已经渐渐消失。现在已经很难区分来自不同地域的居民。或许，可以乐观地认为，我们用文化和视觉上的多样性换取了更为多样的个人生活或者更加多彩、有趣的知识和科技生活。

两个世纪之前丰富多彩的地方生活，通过这幅图册上的图画重现在我们的面前，我们 Manning 的同事庆祝这种发明、创意，并分享着这种用这些图画做计算机图书封面的乐趣。

目录

contents

前言	I
序	III
致谢	V
关于本书	VII
第 1 部分 面向可写的 Web 编程	1
第 0 章 你要先具备哪些知识	3
0.1 你需要了解哪些关于 Java 或 C# 的知识	4
0.2 你需要了解哪些有关 Web 开发方面的知识	5
0.3 你需要了解哪些有关 XML 的知识	6
0.4 博客技术术语	7
0.5 我们要使用的组件	8
0.6 本书是如何组织的	10
0.7 Blogapps 实例	14
0.8 小结	15
第 1 章 全新的协作方式	16
1.1 使用博客进行研究	17
1.2 使用博客发布工作报告	20
1.3 使用博客通知软件构建	21
1.4 博客的商业应用	22

1.5 Nina 和 Rangu 的宏伟计划	25
1.6 小结	27
第 2 章 开发起步	28
2.1 安装博客服务器	29
2.2 开发 Blog Poster 实例	31
2.3 Java 版本的 Blog Poster	32
2.4 C#版本的 Blog Poster	35
2.5 小结	39
第 3 章 走到幕后	40
3.1 博客服务器的结构	41
3.2 wiki 服务器的结构	49
3.3 选择博客还是选择 wiki 服务器	52
3.4 小结	55
第 4 章 Newsfeed 格式	56
4.1 RSS 的诞生	57
4.2 RDF 分支: RSS 1.0	61
4.3 简单分支: RSS 2.0	65
4.4 RSS 的 9 个不兼容的版本	68
4.5 新的标准: Atom	70
4.6 小结	77
第 5 章 如何解析 newsfeed	79
5.1 可行性	80
5.2 使用 XML 解析器解析 newsfeed	81
5.3 使用 newsfeed 库进行解析	91
5.4 开发 newsfeed 解析器	97
5.5 有效地获取 newsfeed	104
5.6 小结	108
第 6 章 Windows RSS 平台	109
6.1 Windows RSS 平台概述	110
6.2 使用通用 Feed 列表管理订阅	117
6.3 使用 Feeds API 解析 newsfeed	124
6.4 Windows RSS 平台的 newsfeed 扩展	130
6.5 小结	139

第 7 章 ROME newsfeed 工具	140
7.1 ROME 介绍	141
7.2 使用 ROME 解析 newsfeed	148
7.3 使用 ROME 获取 newsfeed	158
7.4 使用 ROME 生成 newsfeed	161
7.5 扩展 ROME	163
7.6 小结	176
第 8 章 如何提供 newsfeed	177
8.1 可行性	178
8.2 基本问题	179
8.3 文件库实例	185
8.4 使用 Java 生成 newsfeed	186
8.5 使用 C#生成 newsfeed	192
8.6 有效地提供 newsfeed	197
8.7 小结	205
第 9 章 使用基于 XML-RPC 的 API 进行发布	206
9.1 为什么使用 XML-RPC	207
9.2 Blogger API	210
9.3 MetaWeblog API	211
9.4 使用 C#和 XML-RPC 构建博客客户端	213
9.5 使用博客客户端库	224
9.6 小结	225
第 10 章 使用 Atom 进行发布	227
10.1 为什么使用 Atom	228
10.2 Atom 协议是如何工作的	229
10.3 使用 Atom 协议构建博客客户端	240
10.4 小结	246
第 2 部分 博客小应用程序	247
第 11 章 通过聚合建立群组博客	249
11.1 Planet Tool 介绍	250
11.2 配置 Planet Tool	251
11.3 为 Planet Tool 创建模板	253