

Practical Internet Groupware

实用

Internet 群件



O'REILLY®

中国电力出版社

Jon Udell 著

Tim O'Reilly 序

马远 李钟 译

实用 Internet 群件

Jon Udell 著

Tim O'Reilly 序

马远 李钟 译

O'REILLY®

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

O'Reilly & Associates, Inc. 授权中国电力出版社出版

中国电力出版社

图书在版编目 (CIP) 数据

实用 Internet 群件 / (美) 尤戴尔 (Udell, J.) 著; 马远, 李钟译. - 北京: 中国电力出版社, 2001

书名原文: Practical Internet Groupware

ISBN 7-5083-0656-2

I. 实 ... II. ①尤 ... ②马 ... ③李 ... III. 因特网 - 应用软件 IV. TP393.4

中国版本图书馆 CIP 数据核字 (2001) 第 034193 号

北京市版权局著作权合同登记

图字: 01-2001-2078 号

© 2000 by O'Reilly & Associates, Inc.

Simplified Chinese Edition, jointly published by O'Reilly & Associates, Inc. and China Electric Power Press, 2002. Authorized translation of the English edition, 2000 O'Reilly & Associates, 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 & Associates, Inc. 出版 2000。

简体中文版由中国电力出版社出版 2002。英文原版的翻译得到 O'Reilly & Associates, 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 & Associates, 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 / 实用 Internet 群件

书 号 / ISBN 7-5083-0656-2

责任编辑 / 程璐

封面设计 / Edie Freedman, 张健

出版发行 / 中国电力出版社

地 址 / 北京三里河路 6 号 (邮政编码 100044)

经 销 / 全国新华书店

印 刷 / 北京市地矿印刷厂

开 本 / 787 毫米 × 1092 毫米 16 开本 33 印张 487 千字

版 次 / 2003 年 4 月第一版 2003 年 4 月第一次印刷

印 数 / 0001-5000 册

定 价 / 59.00 元 (册)

O'Reilly & Associates 公司介绍

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 & Associates 公司授权中国电力出版社,翻译出版一批该公司久负盛名的英文经典技术专著。

O'Reilly & Associates 公司是世界上在 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 & Associates 一直处于 Internet 发展的最前沿。

许多书店的反馈表明,O'Reilly & Associates 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 & Associates 公司具有深厚的计算机专业背景,这使得 O'Reilly & Associates 形成了一个非常不同于其他出版商的出版方针。O'Reilly & Associates 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 & Associates 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 & Associates 依靠他们及时地推出图书。因为 O'Reilly & Associates 紧密地与计算机业界联系着,所以 O'Reilly & Associates 知道市场上真正需要什么图书。

作者简介

Jon Udell 曾是《BYTE》杂志新媒体的总编、www.byte.com的首席设计师、《BYTE》的 Web Project 专栏撰稿人。现在，作为一位独立的 Web/Internet 顾问和自由撰稿人，他的作品频频出现在各种书籍杂志及在线出版物上。他的个人主页是 <http://udell.roninhouse.com/>。

封面介绍

本书封面上的动物是海豹和海狮。这两种动物有亲缘关系，它们都是鳍脚目的水生哺乳动物。海狮和长毛海豹都是有外耳海豹科的成员。顾名思义，有外耳海豹头部两侧均有外耳，并被耳垂覆盖。其他所有海豹即典型海豹，都没有外耳，只是在耳朵的位置上长有卷曲的小孔。有外耳海豹与典型海豹的另一个显著区别是它们尾鳍的功能。有外耳海豹在陆地上时能够通过尾鳍向前移动，典型海豹则不行，它在陆地上只能滚动、滑行或者扭动身体。尽管海狮和海豹在陆地上很笨拙，在水里却可以用前鳍优雅地游动。鱼和乌贼是海狮和海豹的主食，这些哺乳动物能够潜到水底两千英尺的地方觅食。

长期以来，人们为了获取海豹和海狮的脂肪和毛皮对它们大肆捕杀。现存的海豹只剩下 18 种，海狮仅 4 种，有些品种已经濒临灭绝。令人欣慰的是，它们现在已全部得到了保护。

目录

序	1
---------	---

前言	5
----------	---

第一部分 Internet 群件应用

第一章 conferencing 概述	19
---------------------------	----

什么是 Internet 群件?	20
------------------------	----

共享的与带注释的数据存储	20
--------------------	----

Email 犯了什么错?	21
--------------------	----

群件的梦想与现实	22
----------------	----

conferencing 的定义	23
------------------------	----

为什么 conferencing 如此重要?	25
------------------------------	----

Lotus Notes、BBS 和 NNTP 新闻组群	25
-----------------------------------	----

第二章 公共在线社区	28
------------------	----

随机新闻阅读器的出现	29
------------------	----

专题公共新闻组的活力	30
------------------	----

在线专题组的运作	31
管理在线讨论	33
混合的 Web/NNTP 讨论系统	36
公共讨论的展望	37

第三章 用 NNTP 和 HTML 进行 intranet 合作 38

应用本地新闻组（综述）	40
conferencing 和 Email	42
对读 / 写 Web 服务器的请求	50
在新闻组上聚集 Web 内容	52
HTML 创作策略	58
接近大众	63

第四章 群件用户的信息管理策略 65

理解和应用讨论区域	65
有效地包装消息和讨论	73
运用消息进行调查	78
不要问 IT 部门能为我们做些什么	85

第二部分 群件文档数据库

第五章 作为群件应用工具的文档数据库 89

Internet 软件必要的简单性	90
文档数据库是如何作为群件应用工具的	92
《BYTE》杂志文档数据库的群件特性	93
《BYTE》虚拟发稿室的群件特性	107
《BYTE》公共和专有新闻组的群件特性	109

第六章 文档数据库输入技术 111

文档数据库系统的概述	112
------------------	-----

文档数据库记录的剖析	114
文档数据库表单模板	116
使用文档数据库表单生成器	118
接受和确认文档数据库记录	122
预览文档数据库记录	125
存储文档数据库记录	128
将报表分配给分析员	134
文档数据库输入技术的展望	138
 第七章 文档数据库导航	140
文档数据库导航控制的简介	142
文档数据库维护的艺术和思考	146
实现 HTML 标签索引	148
实现顺序导航	168
导航控制的静态实现	177
运用 Docbase::Indexer	180
文档数据库导航的展望	181
 第八章 组织搜索结果	182
文档数据库的 Web API	183
搜索引擎的 Web API	185
多种引擎，多个文档数据库	186
URL 名字空间和文档标题名字空间	189
当 URL 和文档标题名字空间不充分时	191
一个实现计划	192
生成结果显示	197
插入 SWISH-E	203
插入 Microsoft Index Server	217
文档数据库搜索的总结	226

第三部分 群件应用和服务

第九章 conferencing 应用 229

例 1: 可评论的文档数据库	230
将 XML 存储库转换到可评论的 Web 页面上	234
用于合作评论的文档数据库	244
文档数据库查看应用程序的功能和限制	258
另一个基于 Email 的评论机制	260
例 2: 支持 conferencing 的帮助桌面应用程序	261
帮助桌面的应用和局限	271

第十章 群件 servlet 273

为什么服务器端的 Java 如此高效?	274
Polls servlet	279
GroupCal servlet	283
作为 Web 组件的 GroupCal	291
群件 servlet 的总结	299

第十一章 会员服务 300

Internet 目录服务的情况	301
LDAP 意见一致性	302
基于订阅者的通知系统	303
在 NT Accounts 数据库中与用户和组工作	316
用 LDAP 与用户和组工作	320
会员服务的总结	323

第十二章 认证和授权技术 325

HTTP 基础认证	325
用于 Apache 的基础认证	327

IIS 的基础认证	328
在 Apache 中保护脚本	330
在 IIS 中保护脚本	330
基于属性的访问	331
用于 Apache 的认证脚本	331
用于 IIS 的认证脚本	333
传递式 ISAPI 认证过滤器	333
基于属性的授权脚本	336
运用 Netscape 目录服务器编写认证	337
授权文档数据库查看器	339
运用 cookie 来授权访问	343
设置 cookie	344
输入 cookie 的内容与持续时间	345

第十三章 配置 NNTP 讨论服务器 348

新闻服务器	350
在 Linux 上安装 INN	351
安装 Microsoft NNTP 服务	354
安装 Netscape 的 Collabra 服务器	363
安装 MS Exchange 的 NNTP 服务	372

第四部分 高级 Internet 群件

第十四章 自动化 Internet 组件 379

对象 Web 和 Internet 群件	380
使用 Web API 作为自动接口	381
XMC-RPC: 下一代 Web API	385
用 Web API 监视和测试群件应用	389
聚合和重新包装 Internet 服务	398

第十五章 分布式 HTTP	408
基于 Perl 的本地 Web 服务器	409
网络 dhttp	432
用 dhttp 检索数据	434
复制数据	437
展望 dhttp	447
复制代码	450
安全 dhttp	452
在 Windows 环境中的 dhttp	456
群件和 dhttp	458
 第十六章 后记	 459
当今 Internet 群件的机会	460
讨论的动态性	461
建立 Web 文档数据库	461
Web 和消息组件	462
应用、组件和服务	463
数据原型和对象持续性	463
高级技术	464
下一代 Internet 群件	464
 第五部分 附录	
 附录一 示例软件	 471
 附录二 Internet RFC: 群件综述	 496
 词汇表	 511

序

在 O'Reilly & Associates 公司，我们总是走在潮流前面。80 年代中期，我们就开始出版关于自由软件的书了，这些程序现在已经融入了 Unix 操作系统，我们编写和出版的书在 Perl、sendmail 和 X Window 系统的应用和普及方面起了重大的作用，其中许多程序已经会聚在了 Linux 的旗帜下。

1992 年，我们出版了《The Whole Internet User's Guide and Catalog》，这本书第一次唤起了公众的 Internet 意识。1993 年我们启动 GNN，这是第一个 Internet 门户网站，也是第一家在 Web 上销售广告的公司。

1997 年我们召开了自由软件开发者会议，这一会议使人们广泛接受了开源软件（Open Source）这一概念。于是整个世界突然明白了，原来最重要的计算机工业革命不是来自大公司，而是来自通过 Internet 共享代码的独立开发者松散联盟。

我们总是试图揭示一种存在的差异，即出版界和学术权威们所讲述的，与工作在技术前沿的程序员、管理员、高级用户的实践之间的差异。事实往往是，当我们吹响口哨时这一技术成为主流也就不远了。

我通常把 O'Reilly & Associates 比作 Internet 革命中的 Paul Revere（译注 1）。

译注 1：Paul Revere（1735—1818），美国独立战争时的英雄，他及时通知了亚当斯等领导人，免遭英国人逮捕。

我说这些并非为了夸耀，而是为了让你相信：当我说到在网上又捞了一条大鱼时，我不是在开玩笑。

偶尔会碰到这样的书，它能使人茅塞顿开，令人叫绝。Jon Udell 的《实用 Internet 群件》就是这样的一本书。

它之所以如此优秀，有如下几点原因：

首先，它有一个清晰的主题：如何为知识的合作管理创造工具。当初次面对 Internet 时的激动和兴奋渐渐隐去的时候，我们想让它提供更多的服务。几乎每一个上网的人都曾被这样的问题困扰：淹没在大量的 Email 中；对着成千上万的搜索结果手足无措；硬盘总是堆满了保存的文件，却不知道如何与需要它们的人共享。

Jon 这本书就是针对以上难题给你实际的指导：使用现代 Web 浏览器中常被人忽略的特性，把有点混乱的 Email 和 conferencing 信息流用 Web 页面组织起来。尽管书中的许多内容是面向开发者的，实际上在商业环境中应用 Internet 的人都能从本书的前几章获益。

如何建立有效的 conferencing 应用程序和其他形式的 Internet 合作是开发者当前面临的重要问题之一。任何人如果想建立有效的 intranet，或者更好地实现公司与客户的交流，或者实现能将人们召集在一起的新方式，那么看完本书后就会有一种全新的概念。

其次，据我所知，Jon 挣脱了封闭我们思维近二十年的桌面计算模式的桎梏。他工作在以 Net、而不是以特定的操作系统为应用开发平台的世界中。

我们经常看到，人们会在袖口（或者 T 恤）上佩带他们的技术派系标记，就如古罗马竞技的骑士。无论应用 NT、Linux，还是用 Perl、Java 或是 Visual Basic 编程——都是差异的标志和产生怀疑的基础。Jon 就从这个四分五裂的世界中如巨人般站立了起来，他只有一个软件信念：实际应用决定一切。于是他自由地穿梭在 Windows 和 Linux、Netscape 和 Internet Explorer，Perl、Java 和 JavaScript 之间，将它们穿插在一起，其关键在于充分理解共享 Internet 协议。

任何想在明天的市场中有所作为的开发者都将需要拥有本书中 Jon 介绍的跨平台工具集。

第三点，或许是最重要的一点，即 Jon 触及了自 Web 引入以来，计算机工业中最重大的变革。

特别是本书的后面几章，他给我们描绘了这样一幅景像：Web 站点本身将作为一种可重用的软件组件。这种模式的改变，其内涵是令人震惊的。我自信地断言在将来的几年中，Jon 在本书中展示的方法论将成为亿万商机和新式软件开发的基础（译注 2）。

就如 Bob Dylan（译注 3）所讲：“事情正在发生，但你不知道发生的是什麼，是吧，Jones 先生？”是的，Jon Udell 已经知道了事情的真相，要是你也想知道，那么我想没有比这更好的开端了。

— Tim O'Reilly
O'Reilly & Associates 公司
董事长兼 CEO

译注 2：此书原版出版于 1999 年 10 月。近来 Microsoft.NET 计划和 Sun JavaOne 的推出，有力地验证了 Jon 和 Tim 的远见卓识。

译注 3：Bob Dylan，美国著名摇滚歌手。

前言

几年前作为《BYTE》杂志的一位高级编辑，我回顾了软件的发展，并写了一些关于技术和工业趋势的文章。1995年春天，在我担任《BYTE》网络版的执行编辑时，世界发生着巨大变化。我所做的正是在1995年几乎所有高技术杂志都迫切要做的事：跳上Web的五彩花车。我兴致盎然地着手这一有梦幻色彩的工作，开始时我的注意力集中在寻找一条便捷有效的途径来实现将《BYTE》转换成电子出版物。但在这条通向Web的路上却发生了一件有趣的事，着手工作的几个星期后，我突然明白了一个道理：原来我们的在线刊物不仅仅是出版物，我发现它很快成为基于Internet的群件（groupware）应用，并且我自己也成了这一应用的创始人。

什么是Internet群件呢？我将它定义成四个相关的原则：

- 一种使用标准Internet（Web、邮件、新闻）客户（程序）、服务器（程序）和协议的方法。
- 一种基于Web、邮件、新闻的应用程序来创建、转换、组织、发送、搜索、出版电子文档的方法。
- 一种管理大量文件的方法，这些文件包含一个企业大量智力资本的半结构化数据。
- 一种支持这些活动的Web、邮件、新闻服务的配置方法。

我花了三年时间，一直在编写并重组这些原则，为的是创造更广泛、更多样的群件方案。有些方案使我自己的部门（一个三人小组）能更有效地合作；而另一些方案则将《BYTE》的全部工作人员联系在了一起，虽然它的办公室分散在新罕布什尔、马萨诸塞和加利福尼亚；还有一些方案则连接了《BYTE》工作人员和几百万在线读者，并使这些在线用户之间也能相互合作。

我在梦想，但也知道我不可能建立这样的应用：

- 一天被一万个人使用。
- 跨越多个专用区（zone）。
- 将任何地点的客户与任何地点的服务器连接。
- 使初级用户也具备 HTML（超文本标记语言）的创作能力。
- 管理结构化（structured）和半结构化（semistructured）信息。
- 运行于所有客户或服务器平台。
- 接入分布式后端服务。
- 可跨服务器集群缩放。
- 能用任何语言书写（C、Perl、Java、Unix shell 和 Visual Basic）。

当然，我从来没想到过会很容易实现这些应用。为什么会这样呢？不是因为我是世界上最聪明的家伙，如果是的话，这个故事就不会如此引人入胜了。事实上，答案在 Internet 本身，它的应用程序和数据结构是以简单的构思和协议为基础的。这就是 Internet 出乎意料地盛行的原因，即使一些主宰潮流的大公司，包括最著名的 Microsoft 都始料未及。你可以使这种简单性为你所用，现在本书将为你指点迷津。

本书面向的读者

从整体上说，本书是为工具制作者和信息设计师服务的，他们的目标是为用户创建面向组的信息管理方案。书中有许多编程的例子，大多用 Perl 编写，一部分用 Java 编写，并涉及很多语言及相关背景知识。例如，我经常用 CPAN（Comprehensive Perl Archive Network，综合 Perl 档案网）模块扩充 Perl，我假设你了解（或者正准备去学习）这些知识。我还假设你熟悉以下内容：