

CGI Programming with Perl

第二版
修订本



CGI编程

—使用*Perl*

O'REILLY®

中国电力出版社

Scott Guelich, Shishir Gundavaram 著
& Gunther Birznieks

孙兆林等 译

CGI 编程——使用 Perl

第二版

Scott Guelich, Shishir Gundavaram 著
& *Gunther Birznieks*

孙兆林等 译

O'REILLY®

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

中国电力出版社

图书在版编目 (CIP) 数据

CGI 编程—使用 Perl/ (美) 古里奇 (Guelich, S.) 等编著; 孙兆林等译. - 北京: 中国电力出版社, 2001

书名原文: CGI Programming with Perl, Second Edition

ISBN 7-5083-0624-4

I .P... II .①古... ②孙... III .Perl 语言—程序设计 IV .TP312

中国版本图书馆 CIP 数据核字 (2001) 第 25878 号

北京市版权局著作权合同登记

图字: 01-2001-2083 号

© 2000 by O'Reilly & Associates, Inc.

Simplified Chinese Edition, jointly published by O'Reilly & Associates, Inc. and China Electric Power Press, 2001. 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 2001。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复制。

书 名 / CGI 编程—使用 Perl (第二版)

书 号 / ISBN 7-5083-0624-4

责任编辑 / 蒙虎

封面设计 / Edie Freedman, 张健

出版发行 / 中国电力出版社 (www.infopower.com.cn)

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 30 印张 447 千字

版 次 / 2001 年 6 月第一版 2001 年 6 月第一次印刷

印 数 / 0001-5000 册

定 价 / 55.00 元 (册)

JS636106

作者简介

Scott Guelich 1993年毕业于Oberlin学院，获哲学学位，然后他决定在继续深造之前休息几年。他没能在现存的哲学流派中找到自己的方向，加上有一些编程经验，他很快转入了计算机行业。此后两年他相继接触到了Internet和Perl。

Scott过去几年一直从事Web开发工作，现在供职于旧金山湾区，他喜欢太极拳、山地车、冲浪、滑雪和任何户外接近大自然的运动。工作之外，Scott是现代技术的抵制者，没有电视，没有投币电话。他仍然希望有一天进入研究生院继续深造。

Shishir Gundavaram 毕业于波士顿大学，获生物制药工程学士学位。在他的毕业论文中，他为NeuroMuscular Research Center（神经肌肉研究中心）的Motor Unit实验室开发了一个Windows程序，用于获取和分析患者的肌肉力量输出，从而间接观察肌肉的电气行为。他是1996年O'Reilly公司出版的《CGI Programming on the World Wide Web》的唯一作者。

Gunther Birznies 是eXtropia.com公司的高级技术干事，该公司以开源Web编程资料与Web编程（如Perl、CGI、Java）有关的在线教程而闻名于世。此前，Gunther在为闻名世界的人类基因组工程做Web编程和信息基础设施建设工作。最近，他成为全球领先的Web工程公司Barclays Capital公司的副主管。

封面介绍

本书封面上的动物是一只老鼠，Muridae科的啮齿动物。纯种（长尾）的老鼠是动物王国最年轻的成员之一，大约出现于1500万年前。现存的老鼠超过200种，最常见的是家鼠。家鼠是地球上分布第二广的哺乳动物，仅次于人。它们通常栖息在人类居住区附近，几乎什么都吃，当然它们最喜欢谷类。

老鼠是公认的有害动物。它们会严重地破坏庄稼，污染食物，还会传播各种疾病。尽管如此，在古代的小亚细亚和希腊，老鼠还是人类的崇拜对象。就是今天，它们在人类文化中的地位也非常重要，经常成为儿童图书和卡通中的英雄，最著名的当然应该算家喻户晓的米老鼠了。

目录

前言	1
第一章 绪论	9
历史	9
CGI 简介	11
可选技术	16
Web 服务器的配置	19
第二章 HTTP	26
URL	27
HTTP	32
浏览器请求	36
服务器响应	44
代理服务器	49
内容协商	51
总结	53

第三章 CGI	54
CGI 环境	55
环境变量	57
CGI 输出	63
示例	73
 第四章 表单和 CGI	 78
发送数据到服务器	79
表单标签	81
解码表单输入	96
 第五章 CGI.pm	 100
概述	101
使用 CGI.pm 处理输入	106
使用 CGI.pm 生成输出	120
生成输出的选择	130
处理错误	134
 第六章 HTML 模板	 141
使用模板的理由	141
SSI (Server Side Include)	143
HTML::Template	154
Embperl	162
Mason	186
 第七章 JavaScript	 189
背景	190
表 单	191
数据交换	203
书签簿	215

第八章 安全性	222
Web 安全的重要性	222
处理用户输入	224
加密	233
Perl 的 taint 模式	236
数据存储	240
小结	243
 第九章 发送电子邮件	 244
安全性	244
电子邮件地址	246
Internet 电子邮件的结构	251
sendmail	252
mailx 和 mail	257
Perl 邮件处理器	258
procmail	259
 第十章 数据保持	 263
文本文件	264
DBM 文件	272
SQL 简介	277
DBI	282
 第十一章 维持状态	 300
查询字符串和附加路径信息	303
隐含字段	310
客户端 cookie	324

第十二章 搜索 Web 服务器	331
顺序搜索引擎	331
两种改进的顺序搜索引擎	335
反索引搜索引擎	340
 第十三章 动态生成图像	 351
文件格式	352
输出图像数据	354
用 GD 生成 PNG	358
附加的 GD 模块	363
PerlMagick	373
 第十四章 中间件和 XML	 381
与其他服务器进行通信	382
XML 介绍	386
DTD	389
编写 XML 语法分析程序	391
从 CGI 网关到 XML 中间件	392
 第十五章 调试 CGI 应用程序	 402
常见错误	402
Perl 编码技巧	406
调试工具	414
 第十六章 CGI 应用程序编程指南	 423
结构指南	423
编码指南	431

第十七章 效率和优化	435
十条基本的 Perl 提示.....	436
FastCGI.....	446
mod_Perl	449
 附录一 参考书籍和相关读物	 455
 附录二 Perl 模块	 460
 词汇表	 465

前言

本书的第一版早在1996年就出版了。那时的网络与现在的网络大不相同：网络主机的数目只有100000，而Netscape Navigator 2.0（第一个支持JavaScript的浏览器）才刚刚发行，同时Java只有不到一年的历史，并且主要用于编写applet（Java小程序）。那时的Web还很年轻，但发展很快。

在1996年，CGI是唯一一种能在网页上生成动态内容的方法，它稳定而且易于理解。然而，很少有网站能充分发挥出它的全部潜能。Shishir在第一版中写到：

当今的计算机用户希望对其感兴趣的问题能有专门的解答。人们满意于计算机中心的工作人员给所有用户分发单一、通用报表的日子已经一去不复返了。取而代之的是，每个售货员、管理者和工程技术人员都想输入特定的查询并且得到及时的回复。如果一台单独的计算机能做到这一点，为什么网络做不到？

这就是CGI的承诺。用户能像专门的工作人员所要求的那样，使用漂亮的饼图或曲线图来按月显示某一特定商品的销售情况，能够使消费者输入关键字来寻找关于商品的信息。

在1996年，这些只是大胆的设想。而今天，它们用来描述商业事务已经是稀松平常的事情。CGI的承诺被完全实现了。

这本书不只涉及到编写 CGI 脚本，它也涉及了 Web 编程。虽然我们着重讨论使用 Perl 进行 CGI 编程（所以这一版改变了书名）这一专题，但我们涉及的许多概念对所有服务器端的 Web 开发是通用的。即使读者将来使用其他技术进行开发工作，现在学习 CGI 也会受益非浅。

本书内容

因为 CGI 在最近的几年里有了巨大的改变，所以反映这些改变就成了本书唯一正确的选择。因此，本书的大部分内容写的都是新主题，包括 CGI.pm、HTML 模板、安全性、JavaScript、XML、搜索引擎、风格建议以及与 CGI 兼容的、高效的替代技术。以前的主题，例如会话变量管理、Email、动态图像和关系数据库，我们进行了扩充和更新。最后，我们更改了 CGI 的叙述方式，以讨论 HTTP（Web 的底层语言）作为开始。对 HTTP 的理解将会有助于更加深入地了解 CGI。

尽管有了这些改变，本书的初衷仍然保持未变：教给读者所需要知道的一切东西，使读者能够成为一个优秀的 CGI 开发者。这不是一本通过范例进行学习的书——它并不是围绕着大量 CGI 脚本进行讨论，并阐述每一个脚本是如何执行的。对于 CGI 来说，已经有很多类似的书可读。这些书当然是很有用的，尤其当其中的某一个范例对应着读者正面临的实际困难的时候更有用，但是，它们通常只告诉读者“怎么做”而没有解释“为什么”。这本书的目的就是：通过阐述基本的原理使读者能够编写 CGI 脚本并且能够解决任何困难。不过，读者不用担心，我们仍将看到很多范例。但是，我们的范例是为了演示阐述过程，而不是为了其他目的。

首先要承认的是这本书侧重于 Unix 平台。Perl 和 CGI 都是最先从 Unix 平台转化而来，因此 Unix 仍然是进行 Perl 和 CGI 开发的最流行的平台。当然，Perl 和 CGI 支持很多其他的平台，其中包括微软公司广为流行的 32 位 Windows 系统：Windows 95、Windows 98、Windows NT 和 Windows 2000（此后统称为 Win32）。就全书而言，我们将着眼于 Unix 平台，但是我们也会指出那些在非 Unix 兼容系统上进行开发所需要注意的问题。

我们使用 Apache Web 服务器实现了全部范例。这有以下几个原因：它是当前最流行的 Web 服务器，可用于大多数平台；另外它是免费的开源软件，并且还支持

许多模块（例如 *mod_perl* 和 *mod_fastcgi*）。这些特性都增强了 Perl 在进行 Web 开发时的能力和性能。

读者需知

读者应该已经适应了 Perl。虽然本书第一版涉及了其他编程语言，但是这一版只集中在 Perl 上。CGI 广泛支持不同的编程语言，但 Perl 肯定是首选。

如果读者还不了解 Perl，那么由 Randal Schwartz 和 Tom Christiansen 编写的《Learning Perl》第二版（O'Reilly 公司出版）将是一个比较好的入门教材（译注 1）。一旦读者掌握了基本知识，我们强烈推荐读者买一本《Programming Perl》第三版（是深受喜爱的骆驼图书），它由 Larry Wall, Tom Christiansen 和 Jon Orwant 编写（O'Reilly 公司出版），是各地 Perl 开发者所使用的权威参考书。在附录一中列出了更多的 Perl 资料。

我们将会论述很多从 *Comprehensive Perl Archive Network*（也就是综合 Perl 文档网络）得来的模块。如果读者从前未曾从 CPAN 下载和安装过模块，请参考附录二。

读者也应该熟悉 *perldoc*，它是浏览 Perl 文件的最好工具。*perldoc* 十分有用，有以下两个理由。第一，它能使用户方便地打开 Perl 编写的大量文件。第二，对于读者学习使用从 CPAN 下载的模块是必需的。*perldoc* 在附录二中有介绍。

本书结构

第一章概述了 CGI，包括它的历史、Web 服务器配置和一个 CGI 脚本的范例。

第二章到第四章讲述了 CGI 的基础。我们从 HTTP 的概述开始着手，然后看看怎样在它上面构建 CGI 程序，最后讲解 HTML 表单，这是传递信息给 CGI 脚本的通用方法。

译注 1：本书中文版《Perl 语言入门》已由中国电力出版社出版。

第五章到第六章介绍了几个常用的模块，可用来帮助我们更加容易地编写 CGI 脚本。也比较了几种不同的生成动态 HTML 输出的技术。

第七章讲述 JavaScript 这种不同的技术如何与 CGI 脚本结合使用，从而生成功能更为强大的方法。

第八章到第十三章给出了 CGI 脚本通常遇到的难题和任务及其解决方法。这些章节包括一些常见主题，例如在线的安全性、存储永久数据和跨页面跟踪用户信息，还包括一些其他的专题，如发送 Email、允许用户搜索你的站点和创建动态图像。

第十四章介绍中间件和 XML，它们使读者的 CGI 脚本能够提供一个到其他信息服务器的接口。

第十五章到十七章通过讲解调试 CGI 脚本的方法以及指导用户编写良好的代码，来阐述怎样编写更好的 CGI 脚本。另外我们还讨论了如何提高系统的性能。

本书还包括两个附录，其中的一个附录给读者提供了更多关于 CGI 信息的资源列表和关于从 CPAN 下载代码的信息。

排版约定

等宽字体 (Constant Width)

用作 HTTP 首部、状态码、MIME 内容类型、配置文件中的指令 (directive)、数组、操作符、变量名 (范例中的除外) 和计算机的文本输出。

斜体 (*Italics*)

用作文件名、路径名、新闻组名、术语第一次介绍时、命令、选项/按钮、程序名、子程序名、函数、方法和主机名。

大写字母 (ALL CAPS)

用作环境变量、HTML 属性和 HTML 标签 (在尖括弧 <> 内的内容)。

建议与评论

本书的内容都经过测试，尽管我们做了最大的努力，但错误和疏忽仍然是在所难免的。如果你发现有什么错误，或者是对将来的版本有什么建议，请通过下面的地址告诉我们：

美国：

O'Reilly & Associates, Inc.
101 Morris Street
Sebastopol, CA 95472

中国：

100080 北京市海淀区知春路 49 号希格玛公寓 B 座 809 室
奥莱理软件（北京）有限公司

询问技术问题或对本书的评论，请发电子邮件到：

info@mail.oreilly.com.cn

如果要寻求技术支持或对本书作出评论，请发电子邮件到：

bookquestions@oreilly.com

我们为本书设立了一个站点，在那里我们列出了范例、勘误表以及对未来版本的一些计划。读者可通过下面这个地址进入这个站点：

<http://www.oreilly.com/catalog/cgi2/>

最后，你可以在 WWW 上找到我们：

<http://www.oreilly.com>

<http://www.oreilly.com.cn>

致谢

因为已经有了写书的经历，我在阅读书中致谢时的感受不同于以前了。一本书凝聚了很多人的大量心血。朋友和家人的贡献更是远远超出了我的想像。

我要感谢我的朋友和家人，他们不仅在我花费大量的时间和精力写这本书的时候给了我很大的理解，一直关心着本书的进度，还耐心地倾听我抱怨闲暇时间太少。还要感谢 Mountain View Printers 公司的员工们，因为这本书很大程度上是在他们夜以继日的工作中才得以完成的。

感谢 HP 公司的 Brad Ashmore 允许我兼职工作，而不是刁难我的工作、写书和折磨我的神智。感谢 Baskar Srinivasan、Natasha Fattedad 和 Anh Hoang 为我填补遗漏。感谢和我在 HP 公司共同工作过的每一位朋友，感谢他们在我觉得不能再继续工作下去的时候所给予我的理解。

我也要感谢在 O'Reilly 工作的每个人。更感谢 Linda Mui 指导这本书出版完成。她总是能解决我的问题，并给出适宜的鼓励和细致的评论。感谢 Rob Romano 所做的插图和 Christien Shangraw 的协助工作。

感谢 Shishir Gundavaram 提供的新材料和我们中很多人都阅读和使用过的初稿。

更要感谢那些审稿人和提供反馈意见的人。Gunther Birznieks 不仅写了一些章节，而且进行了一次非常彻底的审阅。Nat Torkington 进行了更为仔细的审阅。其他给出反馈意见的有 Linda Mui、Andy Oram、Dan Beimborn、Sam Tregar、Paula Ferguson 和 Jon Orwant。

最后，荣誉应归于那些花费很长时间来编写在本书中所讨论的开源应用程序和模块的开发人员。没有他们的工作，网络不会像今天这样精彩。

— Scott Guelich

2000 年 7 月

有很多人参与了这本书的出版。特别是有幸与 Scott Guelich 和 Shishir Gundavaram 这两位我所见过的最有才华的作者共同工作，这真是一次伟大的经历。另外，我要感谢 Andy Oram 和 Linda Mui。通过写作这本书，我从他们那里学到了很多東西。

感谢 Lincoln Stein 最初建议我向 Andy 联系协助出版这本书。也感谢 O'Reilly 提供的其他一些帮助使这本书能够面世。一本书确实是集体智慧结晶。

我也要感谢整个开源社团，为本书的重写提供机会。当我回顾 Perl、Web 技术、人们编写的所有模块以及 Web 全球基础设施的发展时，所有这些已经实现的东西让我大吃一惊。

最后，我想感谢 Electronic Frontier Foundation (<http://www.eff.org/>) 这样的组织，他们致力于保持网络在这个不断发展的法制社会中拥有尽可能多的自由。这种理想建立了一个电脑空间，在这个空间里人们的思想和信息可以自由的飞向世界的每一个角落。

— Gunther Birznieks

2000 年 7 月

第一版的致谢

我要感谢 Dyung Le，他不仅给我提出了写这本书的想法，而且给了我一个离开大学后能够直接开发软件的机会。另外，我要感谢 Rita Horsey 教了我不少东西，并在写这本书的初期提供了 Internet 连接。

当然，我也要感谢我的家人，他们不仅忍受了我在整个写书期间毫无规律的作息時間，而且在任何需要他们的时候给我以帮助。没有他们的支持我无法完成这本书。

感谢所有审稿人和那些提供建议的人：Jeffrey Friedl（他精通正则表达式）、Andreas Koenig（MakeMaker 之父）、Marc Hedlund（CGI FAQ 的创始人）、Tom Christiansen（Unix 天才）、Jon Backstrom、Joseph Radin、Paul DuBois 和来自

O'Reilly 公司的 Norman Walsh、Paula Ferguson、Ellie Cutler、Tanya Herlick、Frank Willison、Andy Oram、Linda Mui 和 Tim O'Reilly（译注 2）。

最后，感谢我所有的朋友、我的家人和我在印度的亲人，尤其是我的祖父母。

希望你们能从这本书中获益。

— Shishir Gundavaram

1996 年 3 月

译注 2：这长长的审稿人名单中可是群英荟萃。J.Friedl 是《Mastering Regular Expressions》一书的作者；A.Koenig 是 C++ 界的名人，名著《Accelerated C++》的作者；T.Christiansen 是《Perl 语言入门》和《Programming Perl》的合作者；P.DuBois 是名著《MySQL》的作者。如此强大的审稿阵容，正是 O'Reilly 公司图书如此成功的原因。