

Solr

实战

[美] Trey Grainger 著
Timothy Potter

范炜 等译

Solr in Action

Yonik Seeley
为本书作序



Solr 实战

[美] **Trey Grainger** 著
Timothy Potter

范炜 等译

Solr in Action

電子工業出版社

Publishing House of Electronics Industry

北京•BEIJING

内 容 简 介

本书介绍了当下最流行的开源搜索技术解决方案 Solr。在搜索引擎视域下，循序渐进地介绍了 Solr 是什么、Solr 能做什么，以及如何更好地使用 Solr 进行开发。在搜索基础层，本书从 Solr 的快速搭建入手，介绍了 Solr 背后的信息检索基本概念，之后重点讲解了构建一个搜索引擎所需的核心模块：索引构建、文本分析、执行搜索及处理搜索结果。在搜索功能层，详细介绍了 Solr 的四大增强型搜索功能：分面搜索、搜索结果高亮、查询建议、搜索结果分组等。在搜索研究的进阶层，介绍了 SolrCloud、多语种搜索及复杂查询操作等。最后，围绕搜索引擎的本质核心问题“相关度”展开了讨论与展望。

本书适合搜索技术工程师、搜索应用设计者以及对搜索引擎技术感兴趣的读者阅读，也可作为高校计算机专业信息技术方向、信息管理与信息系统专业等的课程参考资料。

Original English Language edition published by Manning Publications, USA. Copyright © 2013 by Manning Publications. Simplified Chinese-language edition copyright © 2017 by Publishing House of Electronics Industry. All rights reserved.

本书简体中文版专有出版权由 Manning Publications 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。专有出版权受法律保护。

版权贸易合同登记号 图字：01-2014-4906

图书在版编目 (CIP) 数据

Solr 实战 / (美) 崔·格兰杰 (Trey Grainger), (美) 提摩斯·波特 (Timothy Potter) 著; 范炜等译.
—北京: 电子工业出版社, 2017.5

书名原文: Solr in Action

ISBN 978-7-121-31165-9

I. ①S… II. ①崔… ②提… ③范… III. ①搜索引擎—程序设计 IV. ①TP391.3

中国版本图书馆CIP数据核字(2017)第060465号

策划编辑: 符隆美

责任编辑: 徐津平

印 刷: 三河市良远印务有限公司

装 订: 三河市良远印务有限公司

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

开 本: 787×980 1/16 印张: 39.25 字数: 791 千字

版 次: 2017 年 5 月第 1 版

印 次: 2017 年 5 月第 1 次印刷

定 价: 129.00 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, 88258888。

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

本书咨询联系方式: 010-51260888-819, faq@phei.com.cn。

译者序

搜索无处不在，搜索已经融入我们工作、生活的方方面面。除每天使用的通用搜索引擎以外，几乎我们使用的所有互联网应用、ERP、企业内联网等都提供（或应该具备）了搜索功能。我们向搜索引擎发出查询请求后，很快就能得到一大堆相关的搜索结果，它是怎么做到的呢？这是搜索引擎的黑箱秘密。开源搜索技术 Solr & Lucene 是一套很好的技术工具，能够帮助我们打开这一黑箱，洞悉搜索引擎背后的技术原理与运行机制，并能够借此“真枪实干”地做搜索。

一种观点认为，搜索引擎是信息检索系统的一种类型，即面向网络环境中的海量异构数据的采集、组织与检索。在学科专业领域中，信息检索的研究一直以来都是象牙塔里的一门高深学问。Solr & Lucene 的出现拉近了信息检索学术界与搜索引擎业界之间的距离，加速了搜索应用的设计与开发，新的模型算法与搜索功能在 Solr 的开发技术架构中得以快速转化与实现。

信息检索是“用户需求（行为）—搜索技术（引擎）—信息资源（集合）”三位一体的复合研究主题。搜索研究的原始起点是用户需求与搜索行为，搜索引擎技术是将用户需求与信息资源进行相关度匹配的重要中介手段。没有搜索技术就无从着手开发，不关注用户需求与搜索行为的话，搜索就无从谈起。那么，如何全面理解搜索并开发搜索应用？为了回答这个问题，《Solr 实战》的内容安排试图在技术、资源与用户三者之间找到一种平衡。本书以用户搜索问题解决为导向，通过各种操作实例，全面系统化地介绍了 Solr 的主要功能与使用方法。由于搜索引擎技术体系庞大，涉及数据采集（网络爬虫）、文本处理、自然语言处理、索引、搜索及其他

许多方面，要把这么多理论内容和实践操作放进一本书里，着实不是一件容易的事情，然而本书作者做到了！本书虽然也算得上是一本大部头，但相较于很多大部头技术书，作者对技术细节与技术复杂度的收放考虑还是比较得当的，同时也兼顾涵盖了分布式 SolrCloud、多语种搜索、个性化推荐等热门主题，具有较好的搜索基础普适性与进阶助推性作用。

如果把搜索引擎比作一台汽车，Lucene 与 Solr 的关系就好比是“发动机”与“造车平台”。Lucene 提供了核心的索引与搜索功能，Solr 向下对 Lucene 进行了底层技术封装，隐藏了大量技术细节，降低了进入搜索应用开发的技术门槛。与此同时，Solr 向上扩展了开放集成的大量高级搜索功能，用来满足各种搜索业务需求。因此，Solr 为我们提供了一个更易于学习、操作与应用的全功能开源搜索平台。面对现实的搜索业务需求，从 Solr 入手可以让你进入搜索技术应用层的快车道，通过对 Solr 的深入了解与搜索业务结合，在解决实际搜索问题的大前提下，继而回归到 Lucene 内核的学习。当遇到诸如索引效率问题、相关度模型与算法问题时，自然而然地就深入到 Lucene 内部了。所以，从搜索应用设计与开发的角度来看，从 Solr 入门是一条更有实践效率和学习成就感的途径。

—当前开源搜索技术研究与实践非常活跃，这直接反映在了 Solr 的版本变化上。2010 年 3 月 Solr 开源项目与 Lucene 开源项目合并之后，Solr 成为 Lucene 的一个子项目。从 3.1 版本起，Lucene 与 Solr 的版本保持了一致。版本更新在索引与搜索性能优化、多元化搜索功能引入、交互体验设计、bug 修复等方面不懈地努力着，但 Solr 搜索技术的核心模块并没有发生实质性改变。所以，从搜索技术的学习角度来说，书中使用的 Solr 版本是够用的。Solr 5.0 之后涉及少量基础操作命令的变化，但这不影响书中内容的讲解，我们在书中必要的地方做了注释，方便读者对比。当然，寄希望于读完一本书就能胜任搜索开发恐怕是不现实的。面对网上各种碎片化、不同版本的 Solr 技术资料，对于初涉搜索技术的入门者而言，以这本大部头作为搜索技术学习的小目标，会是一个不错的选择。

本书的初译按章节分工如下：范炜负责第 1、5、6、7、8、16 章，侯任夷负责第 2、3 章，金国栋负责第 4、12 章，邹婧琳负责第 9、10 章，陈皇丹负责第 11、13 章，吴亚平负责第 14、15 章，张功卫、伍志鹏负责附录与其他翻译资料的整理。最终由范炜进行全书的统稿与校对，胡康林参与了校样的审读。

本书的翻译工作历时较长，在此期间王中英与符隆美两位编辑给予了充分信任与理解，并付出了大量辛苦工作，对此表示衷心感谢。鉴于译者专业水平有限，以及专业术语的不一致等问题，译文中难免存在不当之处，恳请读者批评指正。

范炜

2017 年 3 月 4 日于川大江安河畔

原书推荐序

Solr 拥有悠久而成功的历史, Solr 4 和 SolrCloud 开启了全新的篇章。《Solr 实战》的出版非常应景。书中包含清晰的案例、启发性的图表, 涵盖了从核心概念到最新功能的方方面面, 有助于你快速掌握 Solr。

2004 年, CNET 科技资讯网 (现在的 CBS 互动媒体公司) 的搜索引擎服务提供商不再提供服务, CNET 需要一套搜索替代方案, 于是有了 Solr。我开始编写 Solr 的时候, 虽然不具备专业搜索背景, 但做起来很上手。这可能与我的软件设计理念“要快”有关。这一观点有助于 Solr 在传统企业搜索市场以外得以拓展。

截至 2005 年年底, Solr 为 CNET 旗下的许多网站提供了搜索与分面导航功能, 而且很快成为了开源软件。2006 年 1 月, Solr 被捐赠给了 Apache 软件基金会, 成为 Lucene PMC 的子项目, 与 Lucene Java (现在的 Lucene Core) 是兄弟项目关系。由于 Solr 使用 Lucene 作为全文搜索核心库, 两者的技术人员有较大的重合度, 2010 年两个项目就合并在一起了。现在 Lucene 和 Solr 仍然可以分别下载, 但它们的开发由一个团队统一进行管理。Solr 版本号跳跃式地匹配了 Lucene 版本号, 因而两者新版本的发布得以同步。

Solr 4 版本是 Solr 的一个重要里程碑, 增加了 SolrCloud。SolrCloud 拥有一组高度可扩展的功能, 包括无单点故障的分布式索引构建。NoSQL 方面的功能扩展包括事务日志、更新持久性、乐观并发与原子更新。Trey 和 Timothy 作为 Solr 的资深专家和社区成员, 他们撰写的《Solr 实战》一书涉及了 Solr 重要的新功能, 并为 Solr 新手提供了很好的起步指引。

现在 Solr 的应用比我预想的要多得多，涉及图书馆集成系统、电子商务平台、数据分析与商务智能产品、内容管理系统及互联网搜索，等等。Solr 从早期只有少数采用者逐渐成长为大规模的全球社区，在用户的帮助和志愿者的积极推动下，不断向前发展。

《Solr 实战》介绍了 Solr 使用的必备知识和技术，这些都是自 2004 年以来的开发积累。本书在手，可以助你掌握 Solr 的开发。你也可以加入到 Solr 的全球社区，推动 Solr 更好地发展。

Yonik Seeley
Solr 的创造者

前言

2008 年，我受命接管凯业必达（CareerBuilder）招聘网的搜索技术团队。最初使用微软的 FAST 搜索平台，但随后我们意识到，搜索对凯业必达招聘网而言非常重要，与其继续依赖搜索服务提供商，不如在团队内培养搜索专家。我随即开始调研同类型的开源搜索软件，发现 Solr 似乎满足我们搜索开发所需的大部分核心功能。2009 年夏，经过充分的搜索技术准备之后，我们决定将已有的搜索系统转换到 Solr。

Solr 的发展很顺利。Solr 构建在开源搜索库 Lucene 之上。2005 年 2 月，Lucene 正式成为 Apache 顶级项目。2006 年 Solr 被捐赠给 Apache 软件基金会，并于 2007 年 1 月成为 Apache 顶级项目。这两项技术的发展都达到了一定规模之后，在 2010 年 3 月合并为一个项目。

2010 年夏，我们的搜索平台完全转换到了 Solr。在转换过程中，我们提升了搜索速度，大幅度减少了搜索架构所需的服务器数量，避免了昂贵的软件许可费用，增强了平台的稳定性，从先前依赖搜索服务提供商的外部主导转向内部的搜索自主创新。

当时我们并没有意识到搜索自主创新所带来的附加价值。我们已经能够开发出一整套全新的搜索应用产品，包含关键词搜索、语义搜索、大数据分析及时推荐引擎。我们将 Solr 作为可扩展的搜索架构，一小时内通过数百台服务器处理数十亿级的文档与数百万级的查询请求。我们进入了灵活扩展的云服务时代，在数据爆炸中寻求数据的社会意义与价值。Solr 让我们有能力解决迎头而来的挑战。

当 Manning 出版社找到我谈《Solr 实战》一书的写作事宜时，我犹豫了，因为我清楚这是一项艰巨的任务。我提出一个要求：需要一位得力的合作者，而

Timothy Potter 正是合适的人选。Tim 拥有多年 Lucene 和 Solr 搜索方案的开发经验。他利用 Solr 和其他大数据前沿技术，在社交数据的文本分析系统构建与实时分析解决方案方面有着丰富经验。多年来，我俩都得到过 Solr 社区的诸多帮助。当得知开源社区确实需要一本案例驱动的 Solr 指南时，Tim 和我很高兴能撰写《Solr 实战》这本书，为下一代搜索工程师提供帮助。这本书是五年前我们开始接触 Solr 时希望看到的，不论你是刚开始学习 Solr，还是需要补充搜索知识，希望这本书能够对你有帮助。

致谢

与 Solr 一样，这本书的问世离不开广大开源社区成员的支持，在此对他们表示感谢。

- Lucene/Solr 提交者不但能编写令人惊叹的代码，还能提供宝贵的专业知识和建议，同时又能对社区新成员保持足够的耐心。
- 感谢 Lucene/Solr 社区的活跃成员积极贡献代码、更新 Wiki 和其他文档，并回答 Lucene 和 Solr 邮件列表问题。
- 感谢 Solr 的创造者 Yonik Seeley 为本书作推荐序。
- 感谢那些通过 Manning 图书早期访问计划 MEAP 阅读本书草稿，在作者在线论坛上发表评论的读者。
- 感谢在本书写作过程中提供宝贵反馈的评阅人：Alexandre Madurell、Ammar Alrashed、Brandon Harper、Chris Nauroth、Craig Smith、Edward Welker、Gregor Zurowski、John Viviano、Leo Cassarani、Robert Petersen、Scott Anthony、Sopan Shewale 及 Uma Maheshwar Rao Gunuganti。
- 感谢 Ivan Todorović 与 John Guthrie 在本书付印之前的短短时间里对本书进行了详细的技术校对。
- 感谢 Manning 出版社的相关责任编辑们：Elizabeth Lexleigh、Susan Conant、Melinda Rankin、Elizabeth Martin 和 Janet Vail。
- 感谢 Manning 出版社的 Bert Bates 帮助我们提高写作中的讲解说明水平。
- 感谢家人与朋友在我们漫长的研究与写作过程中给予支持。

Trey Grainger

首先要感谢我了不起的妻子 Lindsay，在这本书漫长的写作过程中离不开她的支持与耐心。如果没有她的理解和帮助，这本书不可能完成，特别是在这期间我们的女儿出生了。

我还要感谢 Paula 和 Steven Woolf，他们花了不少时间照看 Melodie，才让我得以顺利完成书稿。最后，我要感谢凯业必达公司团队，包括公司领导层和我的搜索团队，感谢你们让我有机会与优秀的人一起工作，一起构建一个领先的搜索平台，以此惠及社会。

Timothy Potter

我要感谢我的母亲 Sharon Russom，她从小就培养我对学习和书籍的热爱。还要感谢我的父亲 David Potter 对我的学业与工作的支持。没有 Lori Joy 的帮助，这本书不可能完成。感谢你对我在深夜和周末加班的理解与支持，并在写作早期宣传此书。

还要感谢我之前效力的 Dachis 集团。如果没有他们对 Solr 提出那些富有见地的问题，并给我机会去使用 Solr 构建一个大规模搜索解决方案的话，我不可能完成这本书。

关于本书

无论是处理大数据、构建云服务或开发多租户模式的 Web 应用，拥有一套快速且可靠的搜索解决方案至关重要。Apache 的 Solr 是基于 Lucene 的、可扩展的、部署待命的一种开源全文搜索引擎。它提供了开箱即用的多语种关键词搜索、分面搜索、智能匹配、内容聚类和相关度加权等核心搜索功能。

《Solr 实战》是使用 Apache 的 Solr 实现快速且可扩展的搜索的一份权威指南。书中精心准备的案例既有基础的关键词搜索，也有扩展为支持数十亿文档与查询的搜索系统。通过学习本书，你会深入了解如何使用 Solr 的核心功能，例如，搜索结果的分面导航、匹配到的结果片段高亮、字段折叠与搜索结果分组、拼写检查、查询自动补全、函数查询等。你还会了解如何提升 Solr 使用技能，进阶主题包括大规模生产用例、复杂多语种搜索、复杂查询操作及相关度调整的高级策略。

本书框架

本书共分三个部分：“初识 Solr”“Solr 的核心功能”与“Solr 进阶”。对于新接触 Solr 与搜索的一般读者，强烈建议你们按顺序阅读第 1 部分（第 1~6 章），这些章节中出现的许多概念是前后衔接、彼此关联的。

第 2 部分（第 7~12 章）的内容是大多数搜索应用最常见的功能。你可以根据当下的需求关注点，跳过第 2 部分中与需求无关的章节，这样做不会影响其他内容的学习。例如，搜索结果分组是许多搜索引擎的常见功能，但是如果你的数据不需要分组，那么你就可以安全地跳过第 11 章。

第 3 部分（第 13~16 章）的内容有一定难度，介绍了一些高级主题，包括多语种搜索、在大规模集群环境中运行 Solr、高级数据操作及相关度调整等。

书中多数章节都有动手操作，帮助你更好地掌握知识点。我们对每个例子的要求是，既易于操作又能覆盖章节主题。许多例子使用的数据来自真实世界的数据集，以便你接触到现实用例的处理。

第1章介绍 Solr 可以处理的数据类型和应用场景。你会了解 Solr 可以解决的问题类型，从大体上了解 Solr 的主要功能。Solr 4 是 Lucene/Solr 项目的一个重要里程碑，所以即使你对 Solr 先前的版本很熟悉，我们还是建议你阅读第1章，了解一下 Solr 4 带来的全新的、激动人心的功能。

第2章介绍如何在本地工作站上安装和运行 Solr。启动 Solr 之后，我们演示了如何使用 Solr 对一组示例文档集进行索引和搜索。另外，我们还简要介绍了 Solr 的 Web 管理控制台。

第3章介绍了搜索的基础理论，以及 Solr 是如何实践这些理论的。本章最吸引人的内容是倒排索引和相关度评分原理。相关度得分决定了在搜索结果顶部呈现哪些最相关的文档。即使你使用过 Solr，我们还是建议你阅读一下这一章，加深对搜索引擎基本操作的理解。

第4章介绍 Solr 配置的基础知识，主要讲解 Solr 的主配置文件：`solrconfig.xml`。这一章介绍 Solr 最重要的一些配置设定，特别是影响 Solr 如何处理客户端应用请求的那些配置。这一章学到的知识会用在书中其他章节。

第5章介绍 Solr 如何索引文档，从另一个重要的配置文件 `schema.xml` 谈起。你会学习通过定义字段来表示结构化数据，例如，数字、日期、价格和唯一标识符等。这一章还会介绍如何使用 `solrconfig.xml` 对更新请求进行处理和配置。

第6章建立在第5章的内容基础之上，介绍如何使用文本分析索引文本字段。Solr 被设计用于对需要全文搜索的文档进行有效地搜索与排序。文本分析是搜索过程的重要组成部分，它消除了索引文本和查询之间的语言差异。

读到此处，你就已经具备了 Solr 的基础知识，下一步是使用 Solr 来满足你自己的搜索需求。随着对搜索与 Solr 的了解不断加深，你的需求会超越基础的关键词搜索，进而去实现一些常见的搜索功能，例如，高级查询解析、搜索结果高亮、拼写检查、自动建议、分面搜索及搜索结果分组等。

第7章介绍如何构造与执行查询。你会了解 Solr 的许多查询解析器，以及如何对搜索结果进行排序、格式响应、返回及调试等。

第8章介绍 Solr 最强大和最受欢迎的功能之一——分面搜索。Solr 的分面搜索提供了搜索条件限定，通过对搜索结果进行归类，帮助用户发现更多信息。

第9章介绍如何在搜索结果中高亮显示查询词项，以改善搜索解决方案的用户体验。

第10章介绍拼写检查与自动建议。当用户输入几个字符之后，Solr 的自动建议

功能就会给出一个与用户的输入相关的查询建议列表。

第 11 章探讨 Solr 的搜索结果分组与字段折叠功能。当索引包含许多类似的文档时，这样做有助于返回最佳的搜索结果组合，例如，一座城市中同一家餐厅的多个分店。

第 12 章帮助你为在生产环境中部署 Solr 做好准备。这一章会帮助你规划硬件与资源需求，以及决定是否需要考虑使用分片与复制来处理大量的文档与查询请求。

第 13 章介绍 Solr 的分布式功能，也就是 SolrCloud。你会学习如何以云模式运行 Solr，通过对搜索应用进行扩容，以应对大规模用户与文档。读完这一章，你会对 Solr 通过多个服务器上的分布式索引以达到可扩展性与容错性有充分的了解。

第 14 章是第 6 章文本分析内容的延伸，教会你如何在搜索引擎中处理多语种文本。如果需要处理索引中的非英语文本或支持多语种，这一章是必读的。

第 15 章探讨了高级查询功能，包括函数查询、地理空间搜索、多层级分面透视。以及跨文档与跨索引的连接。

第 16 章介绍搜索结果相关度改进技术，例如，调整权重、基于函数的评分、其他相似度算法及相关度得分调试等。此外，进一步讨论了使用 Solr 进行个性化搜索与推荐。

本书有三个附录，对前几章的一些子主题进行了更深入的讨论。附录 A 介绍如何使用 Solr 代码库，为满足特定功能需要和修复官方版本的错误，定制自己的 Solr 分发版本。这个附录是第 12 章开头内容的延伸。

附录 B 以表格形式列出了 Solr 开箱即用的多语种配置。这个附录是第 14 章语种配置内容的扩展。

附录 C 更详细地介绍了数据导入处理器 (Data Import Handler, DIH)，扩展了第 10 章和第 12 章的相关内容，演示了几种大型公开可用数据集的导入步骤。

如何使用本书

《Solr 实战》适合所有软件工程师阅读，无须具备搜索引擎从业经验。本书内容由浅入深，兼顾基础知识与高阶主题，即使经验丰富的 Solr 专业人士也能从最后几章中收获新知。虽然整本书（英文原著）页数超过 600 页，但其中包含了有趣的和实践性的真实案例，在理论与实践之间做出了有效平衡。不论对于刚接触 Solr 还是已有多年 Solr 应用经验的读者，这本书都有一定参考价值。

如前所述，第 1 部分是书中其他部分的基石，这部分内容对 Solr 新手来说非常重要。按顺序阅读这些章节，可以快速有效地掌握 Solr 与搜索技术基础。如果你是 Solr 新手，第 2 章会向你展示如何第一次启动和使用 Solr，第 3 章会介绍搜索的核心理论知识，这些知识是书中其他部分的基础。配置 Solr 服务器并设置字段类型，

以便正确地分析文本内容，这些是掌握 Solr 基本原理的必备知识。

如果你的工作不涉及讨论到的某些功能，可以直接跳过第 2 部分的很多章节。特别是，第 9 章、第 10 章和第 11 章都是比较独立的主题，并不影响阅读后面几章。因此，如果短期内不考虑实现搜索结果高亮、查询建议或搜索结果分组 / 字段折叠，那么可以跳过这三章。第 7 章和第 8 章介绍了许多搜索应用具备的一些常见功能，因此建议你浏览一遍。

其余章节介绍了 Solr 的一些高级主题，解决一些颇有难度的挑战，包括服务器集群扩容、多语种搜索、复杂查询操作以及高级的相关度改进技术。第 2 部分和第 3 部分的章节都基于第 1 部分的内容，第 13 章（SolrCloud）基于第 12 章（搭建 Solr 生产环境），第 15 章（复杂查询操作）基于第 7 章（执行查询和处理搜索结果），第 8 章（分面搜索）与第 16 章（精通相关度）又基于第 15 章。为从本书中获益最大化，请不要跳过作为基础的前几章，它们为后面高级主题的理解提供了必要的知识铺垫。

许多章节包含可执行的示例，你可以一边阅读一边操作。这些示例在介绍新主题的同时，也为你提供了动手探索 Solr 功能的机会。一般来说，你需要做的仅仅是，通过浏览器访问正在运行的 Solr 服务器。你不必运行所有示例，在许多情况下可以直接将它们作为参考配置。不过，运行这些示例会让你获得动手实践经验，这可能有助于理解一些有难度的主题。

是否通读这本书来实现从 Solr 新手到 Solr 专家的进阶，这由你自己决定。如果你暂时不打算通读，当你的兴趣和对 Solr 高级功能的需求增长时，可以随时参考这本书。

代码约定及下载

书中的 Java 代码、配置片段、可执行的命令、文件内容与服务器请求 / 响应（以下统称为“源代码”）均采用等宽字体，以便与周围的普通文本区分开。在许多代码清单中，源代码的关键点都被加以注释。某些源代码被设为粗体等宽字体来进行强调。通过添加换行符和缩进方式来格式化源代码，使其适合书中可用的页面空间。不过，有时还是会遇到很长的代码行，这时会在折行处添加行连续符 ➞。

在整本书中，你会找到 Solr 自带的和书中示例用到的文件来源。文件名除了在源代码中引用时使用斜体外，在其他时候仍使用等宽字体。

书中有很多源代码示例，其中较长的代码清单有明确的标题，较短的代码清单则直接出现在正文中间。书中所有的源代码示例都可以从出版商网站 www.manning.com/SolrInAction 或 www.manning.com/grainger 下载。

源代码的根文件夹中附带一个 README.txt 文件，它详细介绍了如何编译与运行这些示例。由于 Lucene/Solr 项目采用的是 Java 语言，所以我们选择 Java 作为本

书的开发语言。对读者而言，一致的编程语言更容易被接受。

第2章下载并安装 Solr 之后，书中其他部分会使用 `$SOLR_INSTALL` 来表示 Solr 的安装目录。与之类似，`$SOLR_IN_ACTION` 表示源代码下载并解压的文件夹。当看到这两个变量时，请在你的系统里替代成实际的文件夹名称。

作者在线

购买《Solr 实战》可以免费访问 Manning 出版社的内部网络论坛。你可以对本书发表评论、咨询技术问题，并从作者和其他读者那里获得帮助。要访问并订阅该论坛的话，使用浏览器打开 www.manning.com/SolrinAction 或 www.manning.com/grainger。该页面介绍了如何登录论坛、可用的帮助种类及论坛使用规则。

Manning 出版社承诺为读者提供一个场所，方便读者之间、读者与作者之间开展有意义的对话交流。论坛不对作者的参与程度做任何承诺，作者对论坛的贡献是无偿的自愿行为。我们建议读者向作者提出一些具有挑战性的问题，以保持他们有兴趣继续留在论坛上。

本书在版流通期间，从出版商网站上可以访问作者在线论坛和查看之前的讨论归档。

读者服务

轻松注册成为博文视点社区用户 (www.broadview.com.cn)，扫码直达本书页面。

- **提交勘误**：您对书中内容的修改意见可在【提交勘误】处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **与我们交流**：在页面下方【读者评论】处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/31165>



关于封面

《Solr 实战》的封面插图为一名哥特妇女，即来自哥特部落的妇女。哥特人是北方人，2000 年前从斯堪的纳维亚半岛来到欧洲，最初定居在波罗的海。他们曾在罗马帝国的衰落期与中世纪欧洲出现时期扮演了重要角色。他们最终分化成两个分支。西哥特人成为罗马人的联邦，然后西迁到法国和西班牙；东哥特人则迁到意大利北部、巴尔干岛及远东黑海。随着时间推移，他们被定居的地区所同化，自己的语言和文化逐渐消失。

这幅图像摘自 Balthasar Hacquet 的《图说西南及东汪达尔人、伊利里亚人和斯拉夫人》(*Images and Descriptions of Southwestern and Eastern Wenda, Illyrians, and Slavs*)，该书由克罗地亚斯普利特民族博物馆于 2008 年重印出版。Hacquet (1739—1815 年) 是奥地利医生与科学家，在奥地利帝国的许多地方，以及威尼托、朱利安阿尔卑斯山和西巴尔干地区长期从事植物学、地质学和民族志学研究，曾与许多不同部落与民族的人一起生活过。Hacquet 发表的许多论文与出版的书籍中附带了手绘插图。

Hacquet 作品中的图画多种多样，生动地反映了 200 年前阿尔卑斯山和巴尔干地区的独特性与个性。在那个时期，两个村庄仅相隔几英里，根据穿着情况就可以分辨出是哪个村庄的人。民族部落、社会阶层或贸易的成员都可以通过他们的穿着进行辨认。穿着样式和地域特色自那时起发生变化。当时的穿着那么多姿多彩，现如今已经消失了。现在很难区分不同大陆的居民，波罗的海、地中海与黑海沿海城镇村庄的居民与欧洲其他地区的居民没什么太大的区别。

Manning 出版的计算机图书封面采用两个世纪前的生活服装插图（如本书封面插图），旨在宣扬创造性、主动性与趣味性。