

相关性 搜索

利用Solr与Elasticsearch
创建智能应用

Relevant Search: With applications for
Solr and Elasticsearch



[美] Doug Turnbull 著
John Berryman

Trey Grainger 作序

莫映 蔡宇飞 殷智勇 译



相关性搜索

利用Solr与Elasticsearch创建智能应用

Relevant Search : With applications for Solr and Elasticsearch

[美] Doug Turnbull 著
John Berryman

Trey Grainger 作序

莫映 蔡宇飞 殷智勇 译

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书揭开了相关性搜索的神秘面纱，告诉大家如何将 Elasticsearch 与 Solr 这样的搜索引擎作为可编程的相关性框架，从而表达业务排名规则。从这本书中你可学会如何结合各种外部数据源、分类方法以及文本分析手段对相关性进行编程，以满足用户的个性化需求，将令人满意的搜索结果呈现给用户。此外，相关性搜索也需要一定的软性技能。本书还将告诉读者怎样与业务人员协作，为业务找到正确的相关性需求，从而在搜索产品的整个研发生命周期内，实现相关性改进的良性循环。本书介绍了搜索引擎的基本原理，及相关性搜索的调试技术，用大量实例的方式详述了搜索引擎的诸多特性，以形成一整套针对相关性搜索的系统化方法，并倡导致力于提高搜索质量的企业文化。

本书适用于想利用 Elasticsearch 或 Solr 尝试构建智能搜索应用的开发人员。

Original English Language edition published by Manning Publications, USA. Copyright © 2016 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：01-2016-7239

图书在版编目（CIP）数据

相关性搜索：利用Solr与Elasticsearch创建智能应用 / （美）道格·特恩布尔（Doug Turnbull），（美）约翰·贝瑞曼（John Berryman）著；莫映，蔡宇飞，殷智勇译. —北京：电子工业出版社，2017.10

书名原文：Relevant Search: With applications for Solr and Elasticsearch

ISBN 978-7-121-32721-6

I. ①相… II. ①道… ②约… ③莫… ④蔡… ⑤殷… 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228614号

策划编辑：许 艳

责任编辑：刘 舫

印 刷：北京七彩京通数码快印有限公司

装 订：北京七彩京通数码快印有限公司

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：24.5 字数：439千字

版 次：2017年10月第1版

印 次：2017年11月第2次印刷

定 价：99.00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 zltz@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

与本书结缘还要追溯到一年前。当时正值团队启动新产品的研发，需要一款查询性能优良的 NoSQL DB 作为数据存储方案。在考察了包括 Elasticsearch、Solr、Mongo、Cassandra 等一系列 NoSQL DB 之后，我们最终从实际需求出发，选择了与产品功能契合度更高的 Elasticsearch。于是大家开始了对 Elasticsearch 从零起步的探索。不过，在阅读了 Elasticsearch 的大量官方文档之后，大家发现，虽然通过文档的查阅可以了解 Elasticsearch 诸多特性的使用方法，但是这种工具书式的平铺直叙无法将知识有机地联系起来，形成系统而立体的认知。并且，在阅读官方文档的过程中我们也发现，自己对不少搜索相关的基础概念还不甚了解，于是只能借助于网络上搜到的一些支离破碎的快餐资源来补充营养。

为了解决这一问题，大家觉得团队成员们迫切需要一本良师益友式的专业书籍，它既能系统地介绍搜索的相关知识，又能结合当下流行的搜索引擎框架，做到理论与实践相结合。恰巧，电子工业出版社计算机出版分社的许艳老师联系到了我们，商讨一本刚从 Manning 出版社引进的外版书籍的翻译工作，该书正是以 Elasticsearch 和 Solr 为背景介绍相关性搜索的。而且，此书引进时刚刚面世不久，在亚马逊上甚至还没有开始售卖，因此所涉内容的时效性非常之高，正好是团队眼下急需的学习资源。于是，几位同事商量之后，觉得在学习之余，如果顺手将其译成中文，以惠及更多的业内同行，不失为一件利人利己的好事。故而，才有了读者眼前的这本中文版图书。

遇到本书是译者的幸运。书中围绕相关性搜索这一主题，全面系统地介绍了该领域的方方面面：从搜索引擎的基础知识，到相关性搜索的主要技术，再到各种高阶议题，直到当下前沿领域的研究成果，凡此种种，不一而足。两位作者通过朴实细腻的笔触，清晰无误的语言，循序渐进地将我们带入了相关性搜索的神奇世界。这里没有高深莫测的晦涩理论，只有生动有趣的示例讲解。值得一提的是，全书各

章所选的示例多以构建影片搜索应用这一任务为背景，一以贯之，精挑细选。通过来自 The Movie Database (TMDB) 的大量真实影片数据，为读者构建出了一个个实际可运行的搜索示例。其中，以经典系列影片“星际迷航”为主题的影片搜索应用，就在本书的前后多个章节中频频出现，足见作者构思精巧，用心良苦。读完本书，再读 Elasticsearch 或 Solr 的有关文档或书籍时，其中内容无一不有似曾相识的感觉；查阅其他介绍相关性搜索的文章，也有一种“一览众山小”的感觉。而面对现实生活中的各种应用，以及我们自己研发的产品，不禁让人联想，如果将书中所学应用其中，想必一定能画龙点睛，为之增色不少。

本书的翻译过程持续了将近十个月，中文版的字里行间都饱含了几位译者的辛勤汗水。回顾往昔，多艰之旅，历历在目，译稿最终得以成功付梓，实属不易。翻译过程中，从初译到终稿，每一章每一节基本都要经过反复推敲与琢磨至五六遍以上。因为是多人翻译，所以为了保证全书行文风格的统一，最后还进行了一次全面细致的统稿，几位译者都为此投入了极大的精力和时间。大家已经记不清有多少个日夜，当家人都已进入梦乡，自己却还在灯下埋首伏案；有多少个周末，把孩子托付给爱人照料，自己却在一旁奋笔疾书……

本书的翻译也是一次感恩之旅。感谢家人和朋友，没有他们一直以来的支持就不会有本书中文版的问世。感谢莫映的夫人李唯一女士，作为本书的首位读者，每每译稿新鲜出炉，都会经过她的耐心初校，以读者的视角为我们提出诸多中肯的修改建议。感谢智勇的家人，本书翻译之初正值智勇千金呱呱坠地，家人的理解和支持是这位新晋奶爸最大的前进动力。感谢宇飞的爱女，小小年纪就善解人意，能够体谅妈妈因为工作而少了与之相伴的时间。还要感谢博文视点的许艳老师，为我们牵线搭桥，感谢责编刘舫老师，为本书的后期审校尽心尽力。也要感谢我们这几位译者彼此间的相互扶持。大家利用各自的业余时间，以极大的热情投入到翻译工作中，默契配合，一路走来。当然，更应感谢本书的两位原作者 Doug Turnbull 先生和 John Berryman 先生，他们的睿智与经验成就了本书原作的好口碑。

最后，希望中文版的面世，不负原作的美誉，以及各位读者的厚望！

中文版译者

2017年8月于北京，晴耕书斋



蔡宇飞

殷勤

推荐序

在过去十年里，搜索已经变得无处不在——关键字搜索框已经演变成查找数据和浏览大多数网站及应用的事实上的标准用户界面。与此同时，对大多数组织来说，若非被严重忽视，要想提供真正具有相关性的搜索体验一直以来都绝非易事。

强大的开源技术已经能做到在几乎零编码的情况下（如 Apache Solr 和 Elasticsearch），以分布式的、高度可伸缩的方式，实现高效运行和功能丰富的搜索（如 Apache Lucene）。这为几乎所有的开发人员在大数据时代建立起一个“在一般意义上相关（generally relevant）”的实时搜索引擎提供了必要的基础架构。随着搜索在基础架构方面有越来越多的难题得到了解决，加之解决方案的商品化进程，竞争的差异已经从如何提供快速、可伸缩的搜索，转变成如何针对用户的信息需求提供最为相关的匹配。换言之，提供“在一般意义上相关”的结果已经远远不够了——谷歌以及其他顶级的搜索引擎现在已经把用户培养成为这样一种群体，他们期望搜索应用几乎能读懂自己内心的想法。本书所讨论的，就是我们如何更加积极地朝着理解用户意图的方向去努力。

Doug Turnbull 和 John Berryman 是两位经验丰富的搜索和相关性领域的专家，我认识他们已经很多年了，大家时常会在出席搜索大会时遇到。我还能回忆起与他们一起讨论的美好时光，我们共同探讨了如何解决搜索相关性、推荐和个性化方面的一些世界级难题。没有人会比我更加欣喜地看到他们将自己独特的专业知识融入这本书中——这是我读过的十分优秀且引人入胜的技术书籍之一。

相关性调优是一大难题——它经常被误解，而且当出现错误时，往往不会立即显现出来。为了识别出有问题的模式，我们通常需要看到许多错误的样例才行，并且在没有真正看到结果显示的情况下，要想知道什么才是更好的调优结果往往很困难。遗憾的是，通常一个搜索系统直到被部署到生产环境之后，组织才开始意识到

默认提供的相关性和真实的、受领域驱动的、个性化匹配之间存在的差距。

不仅如此，那些处理相关性所需的技能（如领域专长、特征工程、机器学习、本体理论、用户测试、自然语言处理等）与那些构建和维护可扩展的基础架构所需的技能（如分布式系统、数据结构、性能和并发、硬件利用率、网络传输和通信等）也是非常不同的。相关性技术工程师这一角色在许多组织中几乎完全缺失，从而给我们留下了许多未曾发掘的潜力，可以营造出真正让用户满意、并显著推动公司发展的搜索体验。

从手动输入关键字进行搜索到完全自动化的推荐，这一范围内的各种不同程度的个性化，也给我们带来了许多机会，可以为每一位用户的特定需求提供相关性匹配。本书作者相当出色地解释了对搜索特征或信号进行建模时各种方法间的细微差别，从而充分利用了这一范围内的各种不同程度的个性化。掌握了本书中介绍的技术，我们就可以很好地胜任相关性技术工程师的角色，并能够解决在创建真正个性化的相关搜索体验时遇到的许多最具挑战性的问题。

Trey Grainger

Solr in Action 一书的作者

Lucidworks 技术部高级副总裁

前言

John 和我是在共同为 OpenSource Connections (OSC) 做咨询工作、帮助客户解决棘手的搜索问题时认识的。我们有时一起诊断性能问题（好让系统跑得更快），有时帮助构建搜索应用。所有项目都有一系列衡量成功与否的简单指标：系统运行更快了吗？应用开发完成了吗？

但是，搜索相关性并不遵循这些规则。而且从谷歌时代成长起来的用户是不会容忍“还算凑合”这样的搜索的。他们想要的是“绝顶聪明”的搜索。他们希望搜索能够优先考虑其所关心的条件标准，而不是像搜索引擎通常那样，盲目地去猜测相关性。

就像飞蛾被火焰所吸引一样，我们都被这一难题深深吸引了。而且正如所谓的飞蛾那样，我们也常常发现自己是在“扑火”。经过这些惨痛的教训，我们坚持了下来并且得到了成长，在我们最初认为极其困难的任務上取得了成功。

在此期间，我们在 OSC 的博客上也看到了同样的心声。我们意识到有关搜索相关性的问题被记录下来的文字少之又少。于是，我们提出了诸如测试驱动相关性（test driven relevancy）这样的观点。我们记录下了自己心中的困惑、遇到的问题，以及取得的成功。我们一起试验了机器学习的各种方法，比如潜在语义分析（latent semantic analysis）。我们研究了 Lucene 的内部机制并探索了通过构建自定义搜索组件来解决实际问题的相关技术。我们还开始了对信息检索的研究。随着我们所掌握的解决疑难问题的技术越来越多，我们也持续不断地将它们记录为文字。

然而，博客有其自身的局限性。John 和我一直希望以书的形式更加系统地阐明我们的观点。幸运的是，我们经历了一连串有趣的事情，机会往往就会自动找上门来。我在一个本地的技术交流会上与 Andrew Montalenti 一起做了一个关于 Python 并发的演讲。因为 Andrew 在 PyCon 上做过这次演讲，Manning 就给 Andrew 打电话来

讨论写一本关于 Python 并发的书。Andrew 说他对写书不感兴趣，但或许他的联合演讲人 Doug 会感兴趣。

可惜我对写一本关于 Python 并发的书也不感兴趣，但我的确有写另一本书的想法。带着这个想法我找到了 John，经过几番谈话之后，我们共同提出了一个非常激动人心的出书提议——接下来的事大家都知道了！

大约在两年前，我们与 Manning 通了那次意义重大的电话。就像是在坐过山车一样，光阴荏苒，伴随着本书的写作，我们的生活也经历了一系列重大的变迁。我们两个人的家庭都增添了小宝宝。我开启了一项相关性的咨询业务。John 换了工作，成为 Eventbrite 的常驻搜索专家。但我们还是无法拒绝继续为这个让人着迷的话题撰写文章。

你会发现本书不同于其他技术类的书籍，它不是某项技术的功能罗列。它更像是一张地图，指引着我们走出多年的痛苦，去解决那些没有现成答案的难题。换句话说，我们已经走出了搜索相关性的沙漠，发现了许多绿洲，并且学会了如何躲避沙人和帝国突击队。

我们向大家展示这张穿越沙漠的地图，这样大家就不会像我们那样迷失方向。现在，请原谅，我们要找一处最近的海滩来小憩一会儿了……

Doug Turnbull

致谢

在开始撰写本书的几周前，我们两家都迎来了小宝宝。我们要把最诚挚的感谢和爱意送给我们的爱人，Khara Turnbull 和 Kumiko Berryman。我们把连续数个周末的时间都用来写书，而她们也都坚持了下来——在此期间，Khara 完成了她自己的一本书，Kumiko 成功地经历了一次长途越野和房屋出售。现在是时候放个长假了！

本书的成功付梓也离不开 OpenSource Connections 的创始人 Eric Pugh。作为我们的“老板”，是他把我们推到了写作、演讲和解惑的聚光灯下。作为一位领导者，Eric 能够让你的热情成为他的热情。如果不是 Eric 摘掉“辅轮”（有时甚至坚持“独轮”），我们就不会意识到，自己竟然能胜任写作或解惑的工作。Eric 告诉我们，每个人都可以成为思想领袖，包括我们自己。

感谢 TMDb 提供的数据和支持。我们曾经花费大把的时间试图找到理想的数据集。TMDb (<http://themoviedb.org>) 不仅提供了丰富的搜索数据集，而且在我们遇到程序错误和问题的時候（通常是我们自己的代码里的错误），TMDb 也能为我们以及我们的早期读者提供支持。特别要感谢的是 Travis Bell，他总是及时回复我们的问题和邮件。

写书是一项团队活动，我们要感谢 Manning 出版社本书制作团队中的每一位成员：Marina Michaels，我们的开发编辑；Aaron Colcord，技术开发编辑；Valentin Crettaz，技术校对；Frank Pohlmann 和 Mike Stephens，策划编辑；还有负责营销的 Candace Gillhoolley。

我们也要感谢很多参与审稿的朋友，他们阅读了本书最初的书稿，并提出了许多有益的建议，包括 John Guthrie，Martin Beer，Arthur Zubarev，Elman Krinker，Amit Lamba，Marc-Oliver Scheele，Ian Stirk，Joseph Wang，Stuart Woodward，Ursin Stauss，Russ Cam，Michael Fink，Gregor Zurowski，Dimitrios Kouzis-Loukas，

Jeremy Gailor 和 Keith Webster。

另外要感谢 Andrew Montalenti，他为我们与 Manning 建立了联系。还要感谢 Shay Banon 的帮助，他是 Elasticsearch 的创始人，坦率地说，他是一个很和蔼的人。感谢我们的同事，Trey Grainger, Matt Overstreet, Rena Morse, David Smiley, Grant Ingersoll, Yonik Seeley, Rene Kriegler, Peter Dixon-Moses, Charlie Hull 和 Drew Farris，感谢这些年来与我们在搜索和相关性方面的这么多精彩讨论。还要特别感谢 Trey，他为我们这本书写了推荐序。

感谢每一位家人对我们的支持。尤其是我们的孩子们：Megume Berryman, Ian Turnbull 和 Murray Turnbull。感谢我们在 OpenSource Connections 和 Eventbrite 上的“工作大家庭”，让我们能够把大量精力投入到本书的写作上。

关于本书

本书将告诉大家，在响应用户的搜索时要给出用户满意和认可的内容。我们将学习如何根据搜索条件，而不是对搜索引擎的神秘猜测，来严格控制对搜索结果的排名。我们会简要介绍深入定制 Solr 或 Elasticsearch 相关性排名的方法，以及如何采取措施帮助大家发掘相关性对应用而言的意义。

谁应该阅读本书

本书的目标读者是那些渴望了解为什么搜索引擎无法“领会”用户搜索意图的 Solr 或 Elasticsearch 开发人员。对搜索引擎至少有基本了解的读者，可以通过本书将他们的技能提升到更高的层次。虽然这是一本技术方面的书籍，但从组织机构和产品战略的角度来看，它的大部分内容都是围绕相关性展开的，因此也适合于产品经理、内容战略的制订者、市场营销人员，或专注搜索的领域专家阅读。

本书是如何组织的

本书首先介绍了相关技术的基础知识，然后逐步上升到定义和解决搜索相关性问题时我们所要面对的产品策略和文化议题，最后介绍了如何实施个性化搜索、语义搜索以及推荐。

第 1 章从讨论相关性开始。问题涉及的领域包括网络搜索、电子商务，以及专家搜索等。本章讨论了学术界对于我们在相关性领域所做的尝试都提供了什么样的支持。最后，我们简要介绍了本书在解决相关性时所采取的技术策略。

第 2 章对 Lucene 的核心数据结构及其算法做了快速回顾，因为它们与相关性是紧密关联的。我们将会看到，为了寻找相关性内容，基于 Lucene 的搜索为我们提供

了一个如此令人难以置信的框架。

第 3 章告诉大家如何对相关性的进行调试。在第 2 章介绍的数据结构和算法不起作用的时候，我们就需要拿出自己的“工具箱”，弄清楚搜索在哪里出了问题。

第 4 章展示了如何利用搜索引擎的分析流程将内容和搜索分解成可描述的特征。这一基本技巧可以让我们学会如何利用分析手段让所有内容都能被找到。

第 5 章开始讨论针对多个字段的查询策略。在本章中，我们会告诉大家如何构造查询语句，有针对性地去度量那些在搜索阶段对用户而言至关重要的排名因素。

第 6 章继续我们有关查询策略的讨论。在这一章我们重点关注的是以词为中心 (term-centric) 的技术，以及相应的搜索策略，以支持用户对相关性的朴素理解。

第 7 章为大家展示了评价调整 (score-shaping) 技术，比如放大 (boosting) 和过滤 (filtering)。我们时常需要突出近期产生的内容、高利润率的产品，或者距离较近的位置，以此来对搜索加以控制。

第 8 章为大家展示了一系列可以帮助用户找到相关性内容的替代方法。有时，当相关性排名效果不佳的时候，一些 UI 组件，比如可供浏览的切面 (browsable facets)、自动补全 (autocomplete)，以及高亮显示 (highlighting)，也许用这些方法将用户引入正途更为简单。

第 9 章我们构建了一个完整的以相关性为目标的搜索应用，本章将会为大家奉上具有专业眼光的 Yowling。既然已经掌握了相关性技术工程师所具备的一系列技能，本章我们将从头至尾见证一次完整的产品开发流程。

第 10 章从产品战略的角度上升到了一个更高的层面，其目的在于关注文化和组织机构方面的一些因素。一个关注搜索的组织是如何确定何谓相关的呢？我们会看到，一个组织必须实现快速而准确的反馈回路，才能正确引导相关性技术工程师的研发工作。

第 11 章将我们的视野拓展到了搜索引擎以外的地方。本章会向大家介绍机器学习、个性化搜索，以及语义搜索是如何协同工作，一起来提高搜索引擎的相关性排名的。

附录 A 带领大家按照我们曾经走过的流程，利用 The Movie Database (TMDB) API 一步步将本书所用的数据载入 Elasticsearch 中。

附录 B 通过对照 Elasticsearch 和 Solr 之间的相关性功能，指导 Solr 读者阅读本书。

关于代码

本书包含了许多源代码的例子，形式包括带编号的清单，以及普通的文本行。对于这两种情况，源代码都以等宽字体进行了格式化，从而将其与普通文本进行区分。有时代码也会以粗体显示，目的是为了突出相对于本章前面步骤的变化，比如当一个新功能被加入已有的代码行时。

在许多情况下，最初的源代码都已经被重新进行了格式化处理；我们加入了换行并修改了缩进，目的是为了能够适应本书的可用版面。另外，当我们在文中对代码做了解释之后，源代码中的注释通常会被从清单中移除。许多代码清单都会伴有一定的注解，旨在突出显示某些重要的概念。

这些例子都已在 Elasticsearch 2.0 和 Python 2.7 下测试通过。

大家可以在 Manning 的网站 (www.manning.com/books/relevant-search) 上以及本书的 GitHub 库 (<http://github.com/o19s/relevant-search-book>) 中找到第 3 章至第 9 章的代码。为了便于试验，这些例子都是用 iPython Notebook/Jupyter 编写的。README 文件详细说明了运行代码所需的准备工作。

作者在线

购买本书的读者可以免费访问一个由 Manning 出版社运作的私有论坛，在那里你可以对本书发表评论、询问技术问题，并得到作者和其他用户的帮助。要访问和订阅该论坛，请在浏览器中打开 www.manning.com/books/relevant-search。该网页提供的信息包括：如何在成功注册之后加入论坛，你可以得到什么样的帮助，以及论坛内的行为规范。

Manning 出版社承诺为读者提供这样一个场所，在那里不同读者之间，以及读者和作者之间可以建立起有意义的对话。本书作者并不承诺任何具体程度的参与，他们对本书论坛的贡献是自愿的（无偿的）。我们建议大家试着问一些具有挑战性的问题，以激起他们的兴趣！

本书一经出版，就可以通过出版社的网站访问作者在线论坛和以往讨论的存档。

其他在线资源

如果你想了解更多信息，我们推荐以下质量不错的资源。

- OpenSource Connection 的博客 (<http://opensourceconnections.com/blog>)。
- John Berryman 的个人博客 (<http://thoughtbox.solutions>)。
- Elastic 的博客 (www.elastic.co/blog)。
- Lucidwork 的博客 (<https://lucidworks.com/blog>)。
- Salmon Run, Sujit Pal 的 Solr 博客 (<http://sujitpal.blogspot.com/>)。
- Solr Start 的简讯 (www.solr-start.com)。

有关搜索和信息检索方面更为一般性的讨论，我们建议参考下面这部宝典：

- 由 Christopher Manning 等人编写的 *Introduction to Information Retrieval*, (剑桥大学出版社, 2008), <http://nlp.stanford.edu/IR-book/>。

有关 Solr/Elasticsearch 的具体问题，我们建议访问各自的技术论坛：

- <http://discuss.elastic.co>。
- <http://lucene.apache.org/solr/resources.html>。

读者服务

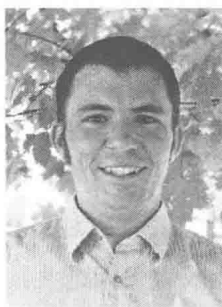
轻松注册成为博文视点社区用户 (www.broadview.com.cn)，扫码直达本书页面。

- 提交勘误：你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（在你购买电子书时，积分可用来抵扣相应金额）。
- 交流互动：在页面下方[读者评论](#)处留下你的疑问或观点，与我们和其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/32721>



关于作者



Doug Turnbull 在 OpenSource Connections 上领导着一项搜索相关性的咨询业务，在那里他经常发表观点和更新博客。Doug 利用各种搜索和自然语言处理技术（NLP）为多个领域的客户构建语义丰富的相关性搜索体验。



John Berryman 的第一份职业是航空工程师，但在航空领域工作了几年之后，他发现编写程序或解决数学难题才是他最喜欢的工作。最终，John 撇下了飞机和卫星，开始全职工作于软件开发、基础架构，以及搜索技术领域。目前，John 供职于 Eventbrite，帮助利用 Elasticsearch 构建事件活动的发现、搜索及推荐。

关于封面插图

本书的封面人物名叫“Homme de l’Isle de Pathmos”，他是来自希腊帕特莫斯岛的一个男人。该图取自 Jacques Grasset de Saint-Sauveur（1757–1810）所著的一本描绘来自不同国家服饰的画册，名为 *Costumes de Différents Pays*，于 1797 年在法国出版。每幅插图都是手工绘制并着色而成的。Grasset de Saint-Sauveur 的这本丰富多彩的画册向我们生动地展示了 200 年前世界各地的城镇和地区在文化上是如此不同。人们彼此之间相互阻隔，说着不同的方言和俚语。在大街小巷或者田间地头，我们很容易就能根据人们的衣着辨别出他们住在什么地方，做些什么买卖，或者身份地位如何。

从那以后，我们的衣着方式发生了改变，那时如此丰富的地域多样性也渐渐消失。现在人们已经很难区分出不同大洲的居民，更不用说不同的城镇、地区或国家了。也许我们已经将文化的多样性换成了一种更加多元化的个人生活——当然是一种更加多元和快节奏的科技生活。

当人们难以分辨不同的计算机书籍时，Manning 根据两个世纪前不同地域的人们在生活上丰富的多样性来设计书籍的封面，以此来颂扬计算机领域里的发明创新和开拓进取，通过 Grasset de Saint-Sauveur 的图片把人们带回到当初的生活。