《干净的数据：数据清洗入门与实践》

图书基本信息

书名：《干净的数据：数据清洗入门与实践》

13位ISBN编号：9787115420475

出版时间：2016-5

作者：[美] Megan Squire

页数：200

译者：任政委

《干净的数据：数据清洗入门与实践》

内容概要

数据清洗是数据挖掘与分析过程中不可缺少的一个环节，但因为数据类型极其复杂，传统的清洗脏数据工作单调乏味且异常辛苦。如果能利用正确的工具和方法，就可以让数据清洗工作事半功倍。
本书从文件格式、数据类型、字符编码等基本概念讲起，通过真实的示例，探讨如何提取和清洗关系型数据库、网页文件和PDF文档中的数据。最后提供了两个真实的项目，让读者将所有数据清洗技术付诸实践，完成整个数据科学过程。
如果你是一位数据科学家，或者从事数据科学工作，哪怕是位新手，只要对数据清洗有兴趣，那么本书就适合你阅读！

《干净的数据：数据清洗入门与实践》

作者简介

作者简介：
Megan Squire
依隆大学计算科学专业教授，主要教授数据库系统、Web开发、数据挖掘和数据科学课程。有二十年的数据收集与清洗经验。她还是FLOSSmole研究项目的领导者，致力于收集与分析数据，以便研究免费软件、自由软件和开源软件的开发。
译者简介：
任政委
辽宁滨城大连现役程序员一枚，长期从事一线软件开发工作，近年来为成为一名“思路清晰”“视角独特”“不搞办公室政治”“输出有生命力代码”“凭借技术知识普惠初中级IT从业者”的终身制全栈式程序员而不懈努力。曾经翻译《Oracle PL/SQL攻略》一书，并希望这本《干净的数据》能够为奋战在IT前线上的各界小伙伴们带来日常工作之外的另类体验。微信号：KNIGHTRCOM

《干净的数据：数据清洗入门与实践》

书籍目录

《干净的数据：数据清洗入门与实践》

精彩短评

1、063.@06142016.新书，逻辑清晰，但浅尝辄止，略失望，也许是我期望太高了.
2、验证一下自己的数据清洗方法。证明数据清晰还是一向初级而琐碎的工作，没有通用的方法。基本上是能用就行。
3、对于外行人来说，貌似需要有一定编程经验；对于相关从业人员来说，太多共识型内容。感觉两个方向都没有做好。作者貌似想从一个又一个的例子告诉你怎么做数据清洗，但是这个东西应该跟特定数据特定需求有关系，不太好一次性说清楚。如果做过多次数据清洗工作的，这本书价值不是很大。

《干净的数据：数据清洗入门与实践》

版权说明