

《数据之美》

图书基本信息

书名：《数据之美》

13位ISBN编号：9787564122720

10位ISBN编号：7564122722

出版时间：2010-6-1

出版社：东南大学出版社

页数：364

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《数据之美》

前言

WHEN WE WERE FIRST APPROACHED WITH THE IDEA OF A FOLLOW-UP TO BEAUTIFUL CODE, THIS TIME about data, we found the idea exciting and very ambitious. Collecting, visualizing, and processing data now touches every professional field and so many aspects of daily life that a great collection would have to be almost unreasonably broad in scope. So we contacted a highly diverse group of people whose work we admired, and were thrilled that so many agreed to contribute. This book is the result, and we hope it captures just how wide-ranging (and beautiful) working with data can be. In it you'll learn about everything from fighting with governments to working with the Mars lander; you'll learn how to use statistics programs, make visualizations, and remix a Radiohead video; you'll see maps, DNA, and something we can only really call "data philosophy."

《数据之美》

内容概要

《数据之美(影印版)》揭示了数据发现可以是多么广泛和美丽！39位业内最佳数据实践者揭秘了他们如何为各种项目开发简单优雅的解决方案。

《数据之美》

作者简介

编者：（美国）西格兰（Toby Segaran）（美国）哈梅巴赫（Jeff Hammerbacher）

《数据之美》

书籍目录

PREFACE 1 SEEING YOUR LIFE IN DATA by Nathan Yau Personal Environmental Impact Report (PEIR)
your.flowingdata (YFD) Personal Data Collection Data Storage Data Processing Data Visualization
The Point How to Participate 2 THE BEAUTIFUL PEOPLE: KEEPING USERS IN MIND WHEN
DESIGNING DATA COLLECTION METHODS by Jonathan Follett and Matthew Holm Introduction:
User Empathy Is the New Black The Project: Surveying Customers About a New Luxury Product Specific
Challenges to Data Collection Designing Our Solution Results and Reflection 3 EMBEDDED IMAGE DATA
PROCESSING ON MARS by J. M. Hughes

章节摘录

插图：We then purchased an inexpensive data set of IP-to-State information. With it, we were able to match each IP address collected with the U.S. state in which it resided. Although we could have scripted our pages to access this database and match the numbers at the time of data collection, we chose to do the matching semi-automatically after the fact. For starters, the project budget and timeframe did not warrant purchasing the additional server power to handle the task. But more important, from a user perspective, was the delay this matching would have inevitably built into the survey completion process. Although it might have been more convenient for us to receive finalized data at once, it would have created an additional inconvenience for our user. When designing a data collection experience, it's important to think about what server tasks must take place during the survey in order for the user's needs to be met, and what tasks can be delayed until after data collection. Don't ask the user to do what you can do—or discover—on your own. All of this leads us back to the central point of this chapter, which is also the final, and core, aspect of building trust: treat the respondent with respect. By demonstrating that you value the respondent and her time and intelligence, by interacting with her in a conversational manner (despite the fact that all survey questions are being delivered by a pre-programmed machine) , and showing her that you've been "listening" to her answers (don't, for example, ask slight variations of the same question over and over again, which makes it seem as though you didn't pay attention to her original response) , you'll increase trust, encourage real answers, and keep the respondent from disengaging.

《数据之美》

媒体关注与评论

数据之美（影印版）“数据实际上已经是下一代计算机应用程序的真正核心。在本书中，业界领先者描述了他们的项目是如何采用新方法来攫取数据的威力。对于那些对数据的未来和解决问题的方法感兴趣的人来说，这是一本必须要读的书。”——Tim O’Reilly, O’Reilly Media, Inc.创始人
和CEO

《数据之美》

精彩短评

- 1、从不同的角度阐述数据挖掘技术应用，难得的好书。
- 2、看着玩的，质量很赞，喜欢
- 3、东南大学出版社出版的东西，还成。中间有彩图。
- 4、这本书适合有一定数据挖掘基础的人看
- 5、内容太杂, 没有目的性.
- 6、借了室友的影印版在读 但是不知道能否坚持 因为缺少相关项目经验 看着那些设计方法没有什么感觉
- 7、翻译的晦涩拙劣
- 8、适合程序，开发，内容图文并茂

章节试读

1、《数据之美》的笔记-1. Seeing Your Life in Data

这是本书的第一章，重点通过介绍Personal Environmental Impact Report(PEIR)和your.flowingdata(YFD)两个应用来阐述对数据从collection到storage, processing, 最终visualization的过程以及在其中每一步的具体设计。

Collection这个步骤中如何让用户最方便的输入数据应该是最重要的了，最好是能够让用户在不知不觉间完成数据的输入，比如PEIR中GPS数据的上传就和用户的行为无关，而YFD借助Twitter这个平台来完成工作也是很好的选择，如果是独立的平台，普通人很难仅仅为了记录和分析自己的日常行为而一遍遍地输入自己所做的事情。

Storage方面介绍了应该根据应用的具体需要设计数据存储方式的思想(有点像废话...)，当然对我的启发是不要做所有事情都直接MySQL...

Processing中用PEIR的数据处理流程为例介绍了一下数据的处理，但总的来说前三个部分在本章当中都没做很多介绍，姑且认为这两个应用在前面三个环节本身就并不复杂。

第一章有一半的笔墨都在介绍Visualization。在PEIR的部分，介绍了其地图显示结合了Modest Maps等几个工具共同完成地图显示的工作。但PEIR这款应用的目的是衡量一个人的交通数据对环境的影响，因此在地图上不仅需要标注出个人的行程路线和起止点等数据，还需要可视化个人对环境的影响，如碳排放等数据。一开始的方案是使用渐变的线条颜色来表示对环境的影响，但这可能与地图上本身的颜色混淆，第二版设计是用终点的圆圈大小来表示影响，但这与正常人的认知不符，人们会觉得这代表着终点发生了什么事情，会迷惑用户。最终的设计方案是将地图颜色反转，整个地图成为深色背景，而路线为浅色，用颜色的渐变表示对环境的影响，而停留点的圆圈大小表示停留的时间，个人认为这个设计也不好，我总觉得深色的背景看久了会让人觉得很压抑。

这一章当中对Data Visualization有一段比较有趣的描写，数据可视化就像是在写一个故事，而如果过多的用图表来表示就会使这个故事看起来像是一本教科书，而用环境，对数据间关系的描述和与用户的互动等方式来可视化数据则像是在写一本小说。

PEIR的数据可视化结果可以分享到Facebook当中，而且它也有Facebook应用，这在如今这个社交网络风行的互联网世界中是很重要的，否则失去了与他人分享的乐趣，我觉得用户可能很难有兴趣坚持下来。

YFD的部分比较少，但我觉得其设计还是不错的，YFD是通过记录和分析用户日常所做的每一件事来形成用户习惯，帮助用户更好的规划和管理生活的应用。首先在界面的中间部分突出了用户的目的，我们做每一件事都是有目的的，因此不但要将事件本身展示出来，更要突出每件事的目的，另一个是对情绪这样一个抽象的概念用标签云的形式进行可视化，用户常用的情绪对应的标签会比较大，这样更为简洁直观，而标签云在当今的互联网应用中还是比较主流的，而且我个人觉得也比较美观。

2、《数据之美》的笔记-

nil

3、《数据之美》的笔记-第33页

这是Chapter 2的笔记，再次吐槽，豆瓣读书的笔记为什么不能选择按章节组织

这一章是以一个电子邮件问卷调查方案的设计过程来讲数据收集的，讲的是UX方面的知识，对专业UX设计人员来说可能讲的都是常识吧，但对不懂这方面的人，例如我，来说是一个很好的案例教学。

摘一些我觉得值得记住的句子吧

Good form design doesn't draw attention to itself and should be nearly invisible " Don't ask the user to do what you can do--or discover--on your ownn " One of the secrets of email surveys is that the second mailing to the same list generally receives just as many responses as the first.类似指数分布的无记忆性？

" you're dealing with extremely thin data sets, so the quality of that data is really important. " In other words, when important decisions are base on the answers given by only a few hundred people, those answers had better be great

4、《数据之美》的笔记-2. The Beautiful People: Keeping Users in Mind When Designing Data Collection Methods

本书的第二章不讲数据，而是讲在收集数据中的人机交互和UX(user experience) design方面的知识，我觉得把这一章拿出来首先讲还是很有讲究的，首先数据收集是第一步，其后才是分析和挖掘，其次用户体验是产品设计和软件开发时最重要的问题之一。

这一章非常浅显易懂，首先讲了UX design的概念，然后用一个新奢侈品调查的案例来具体介绍将数据收集和用户体验结合起来考虑进行产品的设计。首先作为调查乃至用户数据收集的过程，有四个点是一定要考虑的，分别是building trust, length of survey, accurate data collection and motivation。这四点当中除了准确地收集数据外都是从用户的角度来考虑的，目的就是千方百计让用户愉悦地完成调查和数据收集的过程，从而获取较为完整和真实的数据。

下来就开始分析具体的例子了，通过公司前期的调查得出这类奢侈品的主要消费人群是55到75岁的富人，之前的调查方式是通过电话，但这一方式效率很低，人们对于骚扰电话很有戒心，而且电话不一定有人接，即使接了电话也往往是管家、秘书一类的角色，因此采用email的方式，毕竟富人一般还是自己查阅email的。

第一项设计充分考虑到产品面对的老眼昏花的富人，因此采用清楚和平常的字体和较大的按钮，并且在页面布局上注意留白，为了适应多浏览器和多分辨率而采用居中设计等，用户参与这项调查本来就是出于一时兴起，自然需要想方设法顺着用户来设计，不能让其感到一丝不悦。

第二项设计是利用JavaScript动态地显示调查，这样一方面可以控制调查的显示长度，不至于让人望而生畏，另一方面可以根据用户所填选项的不同而在接下来给出不同的调查题目，让整个调查表成为了一个树形结构。但是我觉得这项设计十分欠妥，很多人在参与调查时都希望能提前获知整个调查的长度以建立心理预期，这种设计用于隐藏细节是非常好的，但在本例当中可能会让用户因为不知道多久能完成调查而却步，在调查开头明确标出调查所需时间和题目概数是一个解决方案，不过也不如用户自己浏览一遍题量来得直接有效。

第三项设计是获取信任，表现为在调查开头插入About Us和公司介绍，为了避免引起用户怀疑而采用IP地址来手机用户地址，这一点还是很有必要的，面对55到75岁的中老年人，消除怀疑通常是首先需要做的。

在讲准确地收集数据时提到了在调查表中避免使用劝说技巧来诱导用户购买，因为这项调查的目的是为了研究该产品的市场前景而非进行推广。

下来讲了基于用户动机的设计，这一点比较有意思，调查的头一封邮件设计了两种标题：

" You can shape the face of ... for future generations. "

"We're seeking the voice of experience."

调查表明这些中老年人对于"voice of experience"十分受用，第二种开头的邮件被打开和完成调查的比例都较高，而对于并未理睬第一封邮件的用户，公司又发了第二封邮件，并将两句话都写在了标题中，这次又得到了更高比例的回复(test everything!)

本章最后讲了把收集来的数据通过初步分析汇总成表格的形式提交给奢侈品公司，为了实时性该表格也被放在网页中展示。

5、《数据之美》的笔记-第15页

这是Chapter 1的笔记（话说，我觉得豆瓣的笔记为何一定得按页码组织，做笔记很多时间也需要按章节啊）

这章的作者Nathan Yau 是一个统计学的Phd, 他的博客是<http://nathanyau.com/>

这一章介绍了两个项目：

1. Personal Environmental Impact Report (PEIR)

Nath Yau是这个项目的参与者，这是一个可以让个人了解自己的行为是如何影响环境以及环境如何影响自己的应用。这是个基于位置的应用，利用手机后台中运行的程序，收集同户的地理位置，乘坐交通工具的速度，以及一些环境因素（例如天气），将这些数据代入建立好的环境模型中，最后将个人对环境的影响（如碳排放量）以及环境对个人的影响以数据可视化的形式展现出来。

2. your.flowingdata (YFD)

Nath Yau 是这个项目的创始人，这是个基于Twitter的应用，可以通过收集Tweet来收集的信息，可视化地展现这些信息流，来帮助个人掌握自己的进度之类的，比如减肥，背词之类的。

这一章虽然只有15页，虽然是介绍性的文章，但信息量很大，讲了这两个项目设计中如何将数据收集融入到用户的日常生活中，同步还是异步处理数据的问题，数据存储的方案考虑，数据可视化方案的决策。

其中数据可视化方案的决策，我觉得是这一章最精彩的，受益良多。从可视化方案的总体理念到色彩方案的选择到展现形式的选择。整个过程中都是围绕“ The main design goal behind YFD and PEIR was to make personal data understandable to nonprofessionals ” 来考虑的。

下面摘一段原文，有关数据可视化的理念：I like to think of data visualization as a story. The main character is the user, and we can go two ways. A story of charts and graphs might read a lot like a textbook; however, a story with content, relationships, interactions, patterns, and explanations reads like a novel. This is not to say that one or the other is better. There are plenty of interesting textbooks, and probably just as many-if not more-boring novels. We want something in between the textbook and novel when we visualize personal data.

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu000.com