

<<Collective Intellige>>

图书基本信息

书名：<<Collective Intelligence实战>>

13位ISBN编号：9787302233022

10位ISBN编号：7302233020

出版时间：2010-9

出版时间：清华大学出版社

作者：阿拉克

页数：385

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：<http://www.tushu007.com>

前言

我在2003年4月创立ReadWriteWeb如今已是全球排名前十的技术新闻和分析博客)时,定下的目标是研究当下的Web。

尽管2003年透着.com没落的气息,但还是出现了一些与Web有关的激动人心的新事物。

我之所以将我的新博客命名为Read / Write Web (斜杠和空格已去除),是因为这个新的Web时代似乎体现了Tim Berners-Lee在发明Web时提出的理念,即Web应可被任何人编辑,并且每个人都可以通过某种方式贡献Web数据。

如SatnamAlag在本书中所述,集体智慧(Collective Intelligence)这一研究领域实际上早在Web之前就有了。

但是在.com时代结束之后,我们才开始看到集体智慧运用于Web上的证据。

在2003年,我们不时地会看到它的身影,如在Amazon网站上有用户评论和推荐、在eBay上有用户发起的拍卖、在Wikipedia上有可编辑的百科全书,以及在Google上有用于给网页流行度排名的PageRank算法。

在2004年,O'Reilly&Associates提出了Web 2.0的叫法,最终大多数人认可使用这个词来表示现在这个Web时代(就像.com表示上一个Web时代一样)。

这一新定义的核心部分就是利用集体智慧,将用户贡献聚集起来,以某种方式在网站或应用程序中进行挖掘和利用,这将是很有价值的。

虽然Web 2.0十分流行,但是仍然很难实现其许多原理。

本书可以在这一方面提供帮助,因为它将数学公式和示例应用到集体智慧(从现在起简称为CI)的概念中。

在解释了如何在Web上收集数据和提取智慧后,Satnam在本书第 部分介绍了一些特殊的CI技术,如数据挖掘、文本分析、聚类和预测技术。

<<Collective Intelligence>>

内容概要

在互联网上，利用用户的集体智慧是成功的关键。

集体智慧是一种新兴的编程技术，可让您从人们访问Web和与Web交互的过程中找到有价值的模式、发现这些访问者之间的关系和确定他们的个人偏好及习惯等。

《Collective Intelligence实战》首先介绍了集体智慧的原则和构建更具交互性网站的思想，然后通过示例开发了一个直接可用的基于Java的CI工具包。

您将学会如何从自己的网站和互联网中提取有价值的信息，进而发现流行趋势、做出实际预测和进行推荐。

在此过程中，将使用大量可显著减少开发工作的API和开源工具包。

本书专门为Java Web开发人员而写。

<<Collective Intellige>>

作者简介

作者：（美国）阿拉克（Satnam Alag）译者：腾灵灵 冯飞Satnam Alag目前是NextBio的工程副总裁。他曾是Reartden commerce的CSA（首席软件架构师），拥有加州大学伯克利分校的博士学位。

书籍目录

第 部分 收集数据, 获取智慧 第1章 了解集体智慧(CI) 1.1 什么是集体智慧 1.2 集体智慧在Web应用程序中的应用 1.2.1 通过一个示例全面了解集体智慧 1.2.2 使用集体智慧的好处 1.2.3 集体智慧是Web 2.0的核心 1.2.4 利用CI将以内容为中心的应用程序转化为以用户为中心的应用程序 1.3 对智慧进行分类 1.3.1 显性智慧 1.3.2 隐性智慧 1.3.3 衍生智慧 1.4 小结 1.5 相关资源 第2章 从用户交互中学习 2.1 运用智慧的体系结构 2.1.1 同步和异步服务 2.1.2 事件驱动系统中的实时学习 2.1.3 非事件驱动系统中的轮询 2.1.4 事件驱动和非事件驱动体系结构的优缺点 2.2 应用集体智慧的算法 2.2.1 用户和项目 2.2.2 表示用户信息 2.2.3 基于内容的分析和协作过滤 2.2.4 从非结构化文本中提取智慧 2.2.5 计算相似度 2.2.6 数据集的类型 2.3 用户交互的形式 2.3.1 评分和投票 2.3.2 邮寄或转发链接 2.3.3 书签和保存 2.3.4 购物记录 2.3.5 点击流 2.3.6 评论 2.4 将用户交互转化为集体智慧 2.4.1 一个将评分转化为智慧的示例 2.4.2 来自书签、保存项目、购物记录、链接转发、点击流和评论的智慧 2.5 小结 2.6 相关资源 第3章 从标签中提取智慧 第4章 从内容中提取智慧 第5章 搜索博客圈 第6章 智能Web爬行第 部分 衍生智慧 第7章 数据挖掘: 过程、工具包和标准 第8章 构建文本分析工具包 第9章 通过聚类发现模式 第10章 进行预测第 部分 在应用程序中运用集体智慧 第11章 智能搜索 第12章 构建推荐引擎

章节摘录

插图：随着网站上的内容和商品越来越多，John和Jane越来越觉得手工给商品和其他内容分类是一件繁琐而且耗费时间精力的事情。

同时，用户也反映，内容导航菜单太过生硬。

因此，他们引入了一个新的、动态的导航分类机制：标签云。

在标签云中，标签依照字典顺序排列，且每个标签的字体大小取决于标签的重要程度或者出现次数。

这些标签都是自动地通过对内容的分析后提取出来的。

应用程序分析每个用户的交互，为每个用户提供一组个性化的标签信息作为站点的导航。

如果用户浏览的内容不同，这组个性化的标签也随着改变。

甚至不同的用户在不同的时间点击同样的标签时出现的内容也不同。

这些标签，有的来自于搜索引擎，有的来自于推荐引擎，有的来自于系统外部已有的商品的目录。

在下一个版本中，他们允许用户采用任意的文本作为项目（item）的标签，以及存储或者将有趣的项目保存为书签。

当用户给各种各样的项目加上标签的时候，John和Jane发现这里面含有丰富的信息可以去挖掘。

首先，用户总是用自己觉得有道理的标签去标记项目，这样，他们实际上是在进行公众分类（folksonomy）。

现在，标签云中就不只有原先计算机根据内容生成的标签，也有用户自发产生的标签。

这些用户产生的标签可以组成一个标签词典，用来显著地增强原有的从内容自动提取标签的算法。

不仅如此，这些用户定义的标签还可以当作广告生成系统的关键词，寻找匹配的广告。

这些标签还能把用户和其他用户，或者用户和其他感兴趣的商品联系起来。

群体的智慧正在发挥着效用。

下一步，他们允许用户生成更多的内容。

用户现在可在博客上写下体验，或者在留言板上提出问题和回答问题，还可以向这个网站上的维基贡献维基项目和内容，提升整个应用程序的质量。

John和Jane迅速建立了一种可以从非结构化内容提取标签的算法。

然后，通过对用户交互的分析，他们可以把兴趣类似的用户联系起来，使得用户可以通过其他用户提供的交互信息，发现相关的商品。

他们很快就能够充分了解用户，从而能够提供给每个用户个性化的网站内容，并提供有关内容，比如把小众商品定向推送给小众用户。

通过用户的个人资料以及用户的交互，他们还可以推送相关的广告。

媒体关注与评论

“ 本书兼顾内容和理论，更重要的是，本书介绍的内容切实可行。

” ——Taran Rampersand KnowProse.com “ 本书将教会您如何利用想象的力量。

” ——John Tyler UBS Investment : Bank “ 通过阅读本书，可以学到实用的机器学习。

” ——Robi Sen Twin Tecllllologies “ 本书是有关CI技术的经典之作。

我真希望几年前就已拥有它。

” ——Jerome Betnand Elastic Grid LLC “ 我要向社交网站的所有开发人员强烈推荐本书。

” ——Sopan Shewale TWIKI . NET-Enteprise EIKI

<<Collective Intellige>>

编辑推荐

《Collective Intelligence实战》：用于智能搜索、推荐和预测的可承用代码使用Lucene和Nutch的web爬行和文本分析使用WEKA的机器学习如何实现Java Data Mining(JDM)标准

<<Collective Intellige>>

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:<http://www.tushu007.com>