

<<挖掘社交网络>>

图书基本信息

书名：<<挖掘社交网络>>

13位ISBN编号：9787564126865

10位ISBN编号：7564126868

出版时间：2011-5

出版时间：东南大学出版社

作者：Matthew A. Russell

页数：332

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>

<<挖掘社交网络>>

内容概要

Facebook、Twitter和LinkedIn产生了大量的宝贵的社交数据，但是怎样才能找出谁通过社交媒介进行联系？

他们在讨论些什么？

或者他们在哪儿？

《挖掘社交网络(影印版)》这本简洁而且具有操作性的书将为你展示如何回答这些甚至更多的问题。你将学到如何组合社交网络数据、分析技术，如何通过可视化帮助你找到你一直在社交世界中寻找的内容，以及那些你都不知道存在的有用信息。

每个独立章节介绍了在社交网络的不同领域挖掘数据的技术，这些领域包括博客和电子邮件。

你所需要具备的就是一定的编程经验和学习基本的python工具的意愿。

<<挖掘社交网络>>

作者简介

Matthew A.Russell , Digital Reasoning Systems的工程副总裁和Zaffra的负责人，是热爱数据挖掘、开源和网络应用技术的计算机科学家。他是《Dojo : The Definitive Guide》(O'Reilly出版)的作者。

书籍目录

Preface

1. Introduction: Hacking on Twitter Data

Installing Python Development Tools

Collecting and Manipulating Twitter Data

Tinkering with Twitter's API

Frequency Analysis and Lexical Diversity

Visualizing Tweet Graphs

Synthesis: Visualizing Retweets with Protovis

Closing Remarks

2. Microformats: Semantic Markup and Common Sense Collide

XFN and Friends

Exploring Social Connections with XFN

A Breadth-First Crawl of XFN Data

Geocoordinates: A Common Thread for Just About Anything

Wikipedia Articles + Google Maps = Road Trip?

Slicing and Dicing Recipes (for the Health of It)

Collecting Restaurant Reviews

Summary

3. Mailboxes: Oldies but Goodies

mbox: The Quick and Dirty on Unix Mailboxes

mbox + CouchDB = Relaxed Email Analysis

Bulk Loading Documents into CouchDB

Sensible Sorting

Map/Reduce-Inspired Frequency Analysis

Sorting Documents by Value

cotichdb-lucene: Full-Text Indexing and More

Threading Together Conversations

Look Who's Talking

Visualizing Mail "Events" with SIMILE Timeline

Analyzing Your Own Mail Data

The Graph Your (Gmail) Inbox Chrome Extension

Closing Remarks

4. Twitter: Friends, Followers, and Setwise Operations

RESTful and OAuth-Cladded APIs

No, You Can't Have My Password

A Lean, Mean Data-Collecting Machine

A Very Brief Refactor Interlude

Redis: A Data Structures Server

Elementary Set Operations

Souping Up the Machine with Basic Friend/Follower Metrics

Calculating Similarity by Computing Common Friends and Followers

Measuring Influence

Constructing Friendship Graphs

Clique Detection and Analysis

The Infochimps "Strong Links" API

<<挖掘社交网络>>

Interactive 3D.Graph Visualization

Summary

5. Twitter: The Tweet, the Whole Tweet, and Nothing but the Tweet

Pen : Sword :: Tweet : Machine Gun (!?)

Analyzing Tweets (One Entity at a Time)

Tapping (Tim's) Tweets

Who Does Tim Retweet Most Often?

What's Tim's Influence?

How Many of Tim's Tweets Contain Hashtags?

Juxtaposing Latent Social Networks (or #JustinBieber Versus #TeaParty)

What Entities Co-Occur Most Often with #JustinBieber and #TeaParty

Tweets?

On Average, Do #JustinBieber or #TeaParty Tweets Have More Hashtags?

Which Gets Retweeted More Often: #JustinBieber or #TeaParty?

How Much Overlap Exists Between the Entities of #TeaParty and #JustinBieber Tweets?

Visualizing Tons of Tweets

Visualizing Tweets with Tricked-Out Tag Clouds

Visualizing Community Structures in Twitter Search Results

Closing Remarks

6. LinkedIn: Clustering Your Professional Network for Fun (and Profit?)

Motivation for Clustering

Clustering Contacts by Job Title

Standardizing and Counting Job Titles

Common Similarity Metrics for Clustering

A Greedy Approach to Clustering

Hierarchical and k-Means Clustering

Fetching Extended Profile Information

Geographically Clustering Your Network

Mapping Your Professional Network with Google Earth

Mapping Your Professional Network with Dorling Cartograms

Closing Remarks

7. Google Buzz: TF-IDF, Cosine Similarity, and Collocations

Buzz = Twitter + Blogs (???)

Data Hacking with NLTK

Text Mining Fundamentals

A Whiz-Bang Introduction to TF-IDF

Querying Buzz Data with TF-IDF

Finding Similar Documents

The Theory Behind Vector Space Models and Cosine Similarity

Clustering Posts with Cosine Similarity

Visualizing Similarity with Graph Visualizations

Buzzing on Bigrams

<<挖掘社交网络>>

How the Collocation Sausage Is Made: Contingency Tables and Scoring

Functions

Tapping into Your Gmail

Accessing Gmail with OAuth

Fetching and Parsing Email Messages

Before You Go Off and Try to Build a Search Engine...

Closing Remarks

8. Blogs et al.: Natural Language Processing (and Beyond)

NLP: A Pareto-Like Introduction

Syntax and Semantics

A Brief Thought Exercise

A Typical NLP Pipeline with NLTK

Sentence Detection in Blogs with NLTK

Summarizing Documents

Analysis of Luhn's Summarization Algorithm

Entity-Centric Analysis: A Deeper Understanding of the Data

Quality of Analytics

Closing Remarks

9. Facebook: The All-in-One Wonder

Tapping into Your Social Network Data

From Zero to Access Token in Under 10 Minutes

Facebook's Query APIs

Visualizing Facebook Data

Visualizing Your Entire Social Network

Visualizing Mutual Friendships Within Groups

Where Have My Friends All Gone? (A Data-Driven Game)

Visualizing Wall Data As a (Rotating) Tag Cloud

Closing Remarks

10. The Semantic Web: A Cocktail Discussion

An Evolutionary Revolution?

Man Cannot Live on Facts Alone

Open-World Versus Closed-World Assumptions

Inferencing About an Open World with FuXi

Hope

Index

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>