大数据与人文社会科学研究丛书

Big Data in Humanities and Social Sciences

Digital Humanities

数字人文

信息时代背景下，依托数字化信息开展人文研究的新范式

Big Historical Data

历史大数据

依托大数据探索人类和世界的前世今生，开启历史学家的宏观研究视角

Topic Modeling

主题建模

重构历史数据主体的多样化主题，提供阐释与分析历史数据的新方式

Visualization

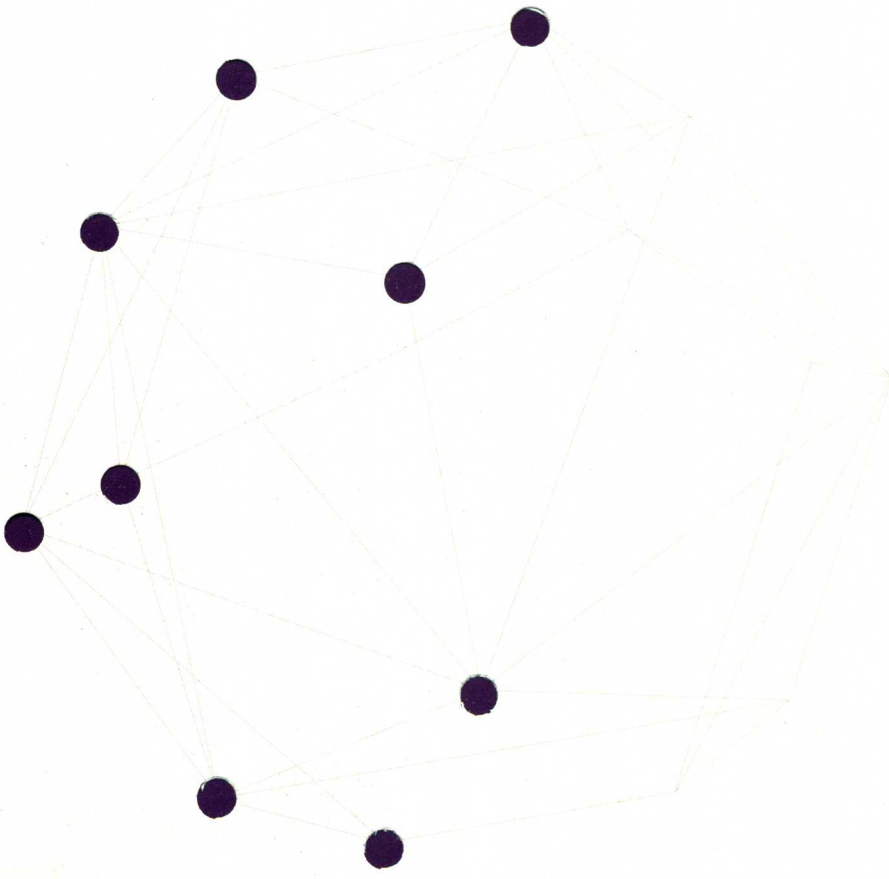
可视化

实现数据的图形化呈现，强化数据交互，助力数据审视

Network Analysis

网络分析

聚焦大数据中的实体及其关联关系，融合微观和宏观研究视角



Exploring Big Historical Data

The Historian's Macroscope

探索历史大数据

历史学家的宏观视角

[加] 肖恩·格雷厄姆（Shawn Graham） [加] 伊恩·米利根（Ian Milligan）
[美] 斯科特·魏因加特（Scott Weingart） 著
梁君英 刘益光 黄星源 译

ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

大数据与人文社会科学研究丛书
Big Data in Humanities and Social Sciences

Exploring Big Historical Data

The Historian's Macroscope

探索历史大数据

历史学家的宏观视角

[加] 肖恩·格雷厄姆（Shawn Graham） [加] 伊恩·米利根（Ian Milligan）
[美] 斯科特·魏因加特（Scott Weingart） 著
梁君英 刘益光 黄星源 译

ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

图书在版编目(CIP)数据

探索历史大数据:历史学家的宏观视角 /(加)肖恩·格雷厄姆(Shawn Graham),(加)伊恩·米利根(Ian Milligan),(美)斯科特·魏因加特(Scott Weingart)著.梁君英,刘益光,黄星源译.—杭州:浙江大学出版社,2019.1

书名原文:Exploring Big Historical Data

ISBN 978-7-308-18858-6

Ⅰ.①探… Ⅱ.①肖… ②伊… ③斯… ④梁… ⑤刘… ⑥黄… Ⅲ.①史学—研究 Ⅳ.①K0

中国版本图书馆 CIP 数据核字(2018)第 302379 号

浙江省版权局著作权合同登记图字:11-2018-572 号

探索历史大数据:历史学家的宏观视角

[加]肖恩·格雷厄姆(Shawn Graham) [加]伊恩·米利根(Ian Milligan)
[美]斯科特·魏因加特(Scott Weingart) 著
梁君英 刘益光 黄星源 译

丛书主持 陈佩钰
责任编辑 陈思佳
责任校对 刘 郡
封面设计 程 晨
出版发行 浙江大学出版社
(杭州市天目山路 148 号 邮政编码 310007)
(网址:http://www.zjupress.com)
排 版 杭州隆盛图文制作有限公司
印 刷 浙江海虹彩色印务有限公司
开 本 710mm×1000mm 1/16
印 张 17.5
字 数 285 千
版 印 次 2019 年 1 月第 1 版 2019 年 1 月第 1 次印刷
书 号 ISBN 978-7-308-18858-6
定 价 59.00 元

本书受中央高校基本科研业务费专项资金资助

supported by the Fundamental Research Funds for the Central Universities

致　谢

本书的创作对我们而言是一次重大的尝试，同时也是一段宝贵的学习经历。我们要向我们的学生和网络版本的所有读者表示感谢，感谢他们参与讨论并与我们进行公开的合作，共同致力于揭示大数据数字化历史的潜力和风险。同行评审提出了极具建设性的建议，使本书终稿的质量有了显著的提升。同时，我们衷心感谢爱丽丝·奥芬（Alice Oven）、凯瑟琳娜·威吉曼（Catharina Weijman）、简·赛耶斯（Jane Sayers）以及帝国理工学院出版社的其他工作人员，他们帮助我们筹划了整个项目，并完善了全书的行文和内容。

本·马威克（Ben Marwick）对我们的代码给予了评价，且多次对代码进行修改使其更清晰，更有效。我们强烈建议大家通过 https://github.com/benmarwick 查看本的 GitHub 存储库，以获取处理大数据的更多方法。大卫·赫西（David Hussey）、金姆·马丁（Kim Martin）和凯特琳·韦恩-赖特（Kaitlin Wain-Wright）以兼具批判性与深刻性的视角仔细审阅并改进了本书的倒数第二稿。乔·范·埃夫里（Jo Van Every）的写作工作坊及其参与者帮助肖恩的工作顺利起步。米歇尔·莫拉韦克（Michelle Moravec）探讨了什么在 Twitter 上有效，而什么没有，并分享了她自己的研究过程，这让我们学到了很多。以利亚·米克斯（Elijah Meeks）、安德鲁·戈德斯通（Andrew Goldstone）、泰德·安德伍德（Ted Underwood）、本·施密特（Ben Schmidt）和迈克尔·威德纳（Michael Widner）给我们带来了诸多启发。蒂姆·埃文斯（Tim Evans）和爱丽丝·奥芬（Alice Oven）推动了本书的诞生。感谢马萨诸塞州历史学会的慷慨，允许我们获取并分析约翰·亚当斯（John Adams）日记中的珍贵数据。最初审阅我们提案的匿名审稿人促使我们开展更具创造性的工作，感谢他们在百忙之中抽出了时间并进行了深思熟虑。

阿曼达·塞利格曼（Amanda Seligman）将部分网站的完善任务分配给了她的学生，作为他们数字化历史研究生课程的一部分。他们的评价也让我们

受益良多。

在宏观层面，以下人员对我们的作品是否能够发挥功效提出了诸多意见、问题和反馈，没有他们的贡献我们无法完成这项工作：安杰拉·泽斯（Angela Zoss）、大卫·拜格利恩（David Baglien）、丹·肖尔（Dan Shore）、涅雅·达德利（Nieya Dudley）、詹姆斯·沃勒（James Waller）、阿曼达·沃克曼（Amanda Workman）、杰西卡·达西（Jessica Daase）、马库斯·范·格林斯文（Marcus Van Grinsven）、贾丝明（Jasmine）、伊绍拉·利昂（Isaura Leon）、阿拉迪（Arlardy）、克里斯蒂安·威廉（Christian Wilhelm）、查尔斯·梅尔伯格（Charles Mehlberg）、安德鲁·赫丁（Andrew Heding）、泽基亚·琼斯（Zakea Jones）、本杰明·高奇（Benjamin Gautsch）、贝丝·穆德拉夫（Beth Mudlaff）、伊丽莎白·杨（Elizabeth Young）、丹妮尔·阿尔瓦罗（Danielle Alvaro）、托尼·休杰（Tony Hugill）、卡莉（Karlie）、妮科尔·斯潘格勒（Nicole Spangler）、阿什利·卡尔森（Ashley Carlson）、布莱斯林（Bricelyn）、威尔·塔金里蒂斯（Will Tchakirides）、露丝·琼斯（Ruth Jones）、埃里克·加多斯蒂克（Eric Gajdostik）、卡西迪（Cassidy）、约翰娜（Johanna）、迈克尔·克莱默（Michael Kramer）、丽贝卡·格里尔（Rebecca Greer）、乔伊斯·周（Joyce Zhou）、克里斯蒂·博斯（Christi Bose）、戴安娜·莫雷诺（Diana Moreno）、迦勒·麦克丹尼尔（Caleb McDaniel）、马滕·迪林（Marten Düering）、约翰·劳登（John Laudun）、扎克·巴蒂斯特（Zack Batist）、乔纳森·麦夸里（Jonathan McQuarrie）、阿里安娜·丘拉（Arianna Ciula）、彼得·霍尔兹沃思（Peter Holdsworth）、克莱门特·勒瓦卢瓦（Clement Levallois）、吉姆·克利福德（Jim Clifford）、乔斯·伊盖尔图安（José Igartua）、罗伯·布莱兹（Rob Blades）、霍利斯·皮尔斯（Hollis Peirce）以及威廉·丹顿（William Denton）。

最后，肖恩感谢塔玛拉（Tamara）、卡里斯（Carys）和科纳尔（Conall）在他被计算机相关问题困扰时表现出的耐心；伊恩感谢詹妮弗·布利克尼（Jennifer Bleakney）对本书进行的大量编辑和复查工作，以及他在滑铁卢大学的同事允许他借助学生的力量对这些想法进行测试；斯科特感谢家人以及合著者在整个过程的每个阶段都显现出了非凡的耐心。

序　言

xv

某位历史学家坐在书桌旁，打开了台灯。她开始认真阅读一摞18世纪伦敦的庭审记录，这些文件是影印版，且质量不佳，她边看边抄录案文。在工作的时候，她开始注意到用于描述年轻女性囚犯的语言似乎存在一些有趣的规律。“我猜想……”她自言自语道，她求助于Old Bailey Online（一个法律在线数据库）并展开搜索工作。很快，她就拥有了一个包含一千份有关女性囚犯的庭审诉讼资料。她下载完整的副本并将其加载到Voyant Tools（一种文本分析工具）中。没过多久，她得到了文本中关键词、关键词搭配以及它们使用频率随时间变化的图表。她更加确信自己的猜测。她使用MALLET（一种处理文本的Java工具包）开始寻找文本中潜在的语义结构。该算法经过多次探索，结果似乎表明，每篇文本的大部分内容都涉及23个常见主题。

那么，这些主题、这些单词列表意味着什么呢？她开始探究主题和文本之间的关系，并发现了一个话语网络，该网络似乎与国家施加给女囚的道德义务紧密相关。她开始探索网络的形式特征，即哪些词汇、什么想法正在从事繁重的语义提升工作？同时，她在语料库上运行RezoViz工具（Voyant Tools中的一个工具），以提取文档中指定的个人和组织。她开始查阅已经提取完成的社交网络，她能够识别女性和看守、儿童和男性的次级社区，并把注意力集中在一群能够将监狱社区凝聚起来的一小部分人身上。不久，她就对18世纪有关女性审判的话语有了深入的宏观认识，对关键的个体、组织以及它们之间的联系同样了然于心。她看了看表，两个小时过去了。她对这样的结果感到满意，然后停止了此次历史宏观探究，她关掉了电脑，再次将目光投向手边的抄本。

我们生活在一个人文学者需要了解如何利用数字化媒介进行传统人文学术研究的时代，包括它们的算法、假设、用法和中介。人文学科和数字化媒介

xvi

（也称“新媒介”）的联系可以追溯到几十年前，且两者的交互促进了彼此的发展。从广义的视角来看待“新媒介”，我们可以发现，对于先前通信技术的引入以及它们代表（或者说是“构建”）人类知识的方式相应地也需要新的观点并采取新的方法。上文中的例子，为我们展现了一种历史学家利用历史学领域“大数据”进行研究的可能方式。除此之外，还存在其他路径。本书的三位作者已经探索了许多可以用于历史学和其他人文学科研究的大数据工具和研究视角。这些方法正在不断发展、完善，本书挑选了其中最有用的一些方法进行介绍和描述，内容涵盖它们的使用方法、注意事项、各类问题以及宏观研究开创的全新视角。

我们将这本书的副标题命名为“历史学家的宏观视角”，以此表明这既代指一种工具，也代表了一种研究视角。我们并不是在暗示这就是历史学家遇到大数据时“做”历史的唯一方式；相反，它只是工具箱中的一种手段，是历史学家处理无法回避的“大”数据的又一种方式。更重要的是，所谓的“宏观”（一种观察体量巨大事物的工具）恰恰体现了一种科学家的工作平台，研究人员借助不同的工具来探究不同的问题，并在笔记本上留下记录。同样地，（我们认为）历史学家的大数据路径需要的是一种公开的方法，历史学家保留公开的记录，以便其他人可以通过保留的信息探索相同的路径，同时可能会得到完全不同的结论。这是一种“生成性”的方法：人文科学的大数据不仅可以证实过去的故事，而且能够生成新故事、新观点，为我们提供新的工具和优势。

本书结构

本书分为三大部分。在前两章中，我们会对该领域进行总体性概述。第一章将简要介绍大数据时代，以及为什么我们认为这对历史学家至关重要。我们首先探讨大数据对于人文学科研究者的意义，通过关注几个重大项目来探究该领域的现状，然后简要回顾我们自身领域的历史，从那个听上去不太可
xvii 能的起源——罗马教皇的格列高里大学中的一名牧师说起。随后我们会探讨大数据和学会，以及我们为何相信历史学家正在投身计算革命的“第三次浪潮”。第二章将继续概述部分，并将重点放在更为具体的问题上，即什么是数

字人文(简称“DH”)时刻？我们简要介绍了一些关键性术语，包括开放获取、版权和数据挖掘等。我们认为即使历史学家并不把自己定义为数字化历史学家，但他们已经开始以一种温和的方式展示如何构建属于自己的历史学家工具箱。该章结束时将讨论如何获取自身的历史数据，并为后文做铺垫。

第三章开启了本书的第二部分，重点介绍具体的文本分析工具，并对第二章末尾提到的几种用以获取数据的数据挖掘工具进行解释。我们循序渐进地介绍 word clouds(词云)和其他现成的软件是如何帮助你快速把握大规模文本的。随后，本书开始探讨正则表达式。我们在此给你提个醒，学习正则表达式并非易事，这一部分可以被视为对“正则表达式是什么”的介绍及其参考内容。你可以用正则表达式完成令人惊叹的工作，对于任何离不开文本的历史学家来说，正则表达式都是非常有用的工具。

对于第三章中的许多工具而言，你必须明白使用它们想要达到的目的。相应地，在第四章中，我们将对主题建模的多种方法进行深入探讨。这是数字人文学科中最令人兴奋的新工具之一，简而言之，它将重构你正在探索的数据主体或语料库的各种“主题”。这种工具可能有点复杂，所以我们一开始就提供了“手动主题建模”练习，以此帮助你理解。这绝对是一个跟实际操作紧密相关的章节，最终会为你带来巨大的回报和收益。

本书的第三部分由第五章、第六章和第七章组成，主要对网络进行深入的分析。因为网络既是一种分析方式，也是一种强大的可视化形式，我们花了一些时间来讨论可视化的基础知识，试图消除学者对大数据的一个主要担忧：只见森林，不见树木。我们认为，网络分析可以让历史学家将个人参与者置身于复杂的关联关系之中，同时把微观和宏观结合起来。网络分析已经成为主题建模可视化极其有成果的方式之一，这在前一章也有所提及。对于历史学家来说，网络分析有助于他们深入探索空间和时间的概念。第六章会介绍网络分析的概念和词语的基本分类，这是一种独特的革命性建模技术。所有这些 XVIII
都可以借助电子表格程序或其他基础性软件来执行。之后，第七章将探讨更详细的主题和它们带来的机遇，以及更多帮助我们进行网络分析的实用工具。

在结论部分，我们额外指出了一些你可能感兴趣的内容，特别是如何推广你运用本书技能完成的非凡工作。我们还将回顾序言开头提到的那位历史学家，看看她的工作在尝试“宏观研究”之后有了怎样的进展。本书并不是要对

整个数字化历史学领域进行详尽的介绍,任何作品都无法做到这一点。我们是三位扎根北美的学者(其中两位来自加拿大,一位来自美国),而我们举的例子通常来自我们熟悉的学术环境。本书主要关注文本本身。我们努力举出多样化的例子,但我们承认我们的写作起源于北美,且我们更加偏爱英文资料。此外,本书尤其关注文本分析、文本处理以及文本网络,不涉及历史地理信息系统(GIS)或数据库理论。有关GIS的研究已经有了诸多成果,包括最近由卡尔加里大学出版社出版的开源出版物《加拿大历史GIS研究》(*Historical GIS Research in Canada*),该书可从以下网址下载:http://uofcpress.com/books/9781552387085。

本书背景

肖恩(Shawn Graham)一直打算撰写一本有关大数据与历史研究的著作,并在2013年3月得到了帝国理工学院出版社(ICP)的认可。仅从定义来看,有关"大历史"的图书几乎无法由某一位学者独立完成,肖恩很快邀请了伊恩·米利根(Ian Milligan)和斯科特·魏因加特(Scott Weingart)与自己合作。我们三人一起利用Google doc(谷歌共享文档)完成了一份计划书,并将其提交给ICP的委托编辑爱丽丝·奥芬(Alice Oven)。ICP将该计划书发送给四位匿名同行评审,他们快速且极具建设性的反馈使我们能够完善图书的计划书。ICP的董事会在2013年夏季批准了这份计划书。

我们与ICP讨论的一个关键点在于我们希望能以公开的方式开展这项研
xix 究,我们希望进行生活写作和试验,希望我们的工作能够覆盖更广泛的数字化历史社区。显而易见,这样的想法带来了一个巨大的担忧,即可能不利于图书的销售。经过多次讨论,我们最终达成了以下共识:我们将明确声明最终发布的产品在结构和连贯性等方面不同于可能会出现在网络上的版本,并且我们会为实体和电子版本的图书提供网络链接。我们同意如果有数据表明图书销售情况受到了网络版本的影响,会立即下架网络版本。这是合情合理的做法。高质量的编辑指导使该项目优越性突出,但如果该版本存在缺陷,那么你我都不希望你引用本书的观点或把我们的结果作为相关研究的基础。事实上,我

们认为在公开的情况下犯错不是一件坏事，这样我们就能迅速完善作品（这与开源软件发展的关键如出一辙）。多个成功项目带给了我们诸多启发，如多尔蒂（Dougherty）和纳兹多夫斯基（Nazrawtoski）的作品《在数字化时代书写历史》（*Writing History in the Digital Age*），这些项目开启了我们的写作过程，比如肖恩很肯定他的作品由于参与了这种试验而得到了明显改善，我们十分期待接下来会发生什么。[①] 网站会继续发展并对用户做出回应。甚至在初稿完成之前，本书就已经被列入了美国研究生阶段数字化历史课程的教学大纲，我们对此十分高兴。我们鼓励用户为我们的网站（http://www.themacroscope.org/2.0/answers/）提供反馈意见，并承诺会对各种问题、新的进展、问询和评论做出回应。不仅如此，即使网站可能会因为一些不可预知的情况消失，我们也将确保它能通过 archive.org 网站上的 Internet Archive（数字化图书馆）得以保存。

我们很快就完成了整合工作。如何加快工作进度始终是我们最关心的问题，毕竟在这个日新月异的研究领域，人们需要相对迅速地开展工作。我们尽可能地遵从一般性原则，但即便如此，这些内容也时常出现在特定的软件环境或人文学者使用的工具之中。我们不能用五年时间来完成这样一本书。我们必须加快进度，但绝不能匆匆忙忙。

本书的目标读者

XX

我们理想的读者是高年级本科生，他们正处于初次接触大数据并需要相关指导的阶段。为了方便本书用于本科生的数字化历史课程，我们建议教师将本书与《编程历史学家》（*The Programming Historian*）（可从以下网址获取：http://programminghistorian.org）以及丹尼尔·柯恩（Daniel Cohen）和罗伊·罗森茨魏希（Roy Rosenzweig）合著的《数字化历史：利用网络收集、保存和呈现历史指南》（*Digital History: A Guide to Gathering, Preserving,*

① Jack Dougherty and Kristen Nawrotzki (2013). *Writing History in the Digital Age*, Ann Arbor: University of Michigan Press. 该书可从以下网址获取：http://www.press.umich.edu/6589653/writing history in the digital age。

and Presenting the Past on the Web)配合使用。[①] 师生还应该密切关注网站 http://digitalhumanitiesnow.org。我们认为本书还能为以下几类读者提供帮助：需要处理大量数据的研究生、初次接触这些方法的研究人员、那些偶然发现了大量宗谱数据或数字化报纸的有趣之人，以及当地历史学会中试图第一时间获知数字化能够带来多大价值的人。我们致力于把术语的门槛降到最低，并且任何时候都不试图在没有根据的情况下对数字化的精通水平做出假设。的确，我们假设翻开本书的你有兴趣进一步发挥计算机的潜能，而不仅仅是使用在线媒介或者输入一个 Word 文档，而且你希望积极创建和探究数字化数据。

此外，我们假设你愿意真正动手去做。

但有些时候，本书或许会存在一些遗漏，也许有一些东西讲得不够清楚。请你在本书的网站上给我们留下评论，我们会尽力进行补充和完善。另外，你应该了解以下网站。

Digital Humanities Questions and Answers(数字人文问答)：

http://digitalhumanities.org/answers/.

Stack Overflow(堆栈溢出)：

http://stackoverflow.com/.

这两个网站旨在帮助那些需要信息与拥有信息的相关人员建立联系。如果你有任何问题，首先要做的就是直接把问题输入搜索引擎。如果你没有找到满意的答案，请在上述任一网站上注册并提问，把问题问得尽可能具体些：
XXi 确保你提及了正在使用的程序和相关操作系统，并举例说明你为了解决该问题都做了哪些努力。

这么做最棒的地方在于，当你提问时，你其实正在帮助将来可能遇到类似问题的其他人。还有一些网站也会让你眼前一亮。

Digital Humanities Now(当代数字人文)：

http://digitalhumanitiesnow.org.

---

① Daniel J. Cohen and Roy Rosenzweig (2005). *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*, Philadelphia: University of Pennsylvania Press.

The Journal of Digital Humanities(数字人文学报):

http://journalofdigitalhumanities.org/.

上述两个网站是在2015年前后数字人文领域处于领先地位的在线出版物。前者策划并收集活跃的研究人员和编著者的博客,然后经过筛选把一部分收入更正式的《数字人文学报》(*Journal of Digital Humanities*)。这些出版物收藏价值极高,记录了该领域的发展历程。

以下是与该领域有关的Twitter。

丹尼尔·柯恩(Daniel Cohen)完好地保留了列表:https://twitter.com/dancohen/digitalhumanities.

肖恩的Twitter:https://twitter.com/electricarchaeo.

伊恩的Twitter:https://twitter.com/ianmilligan1.

斯科特的Twitter:https://twitter.com/scott_bot.

快去关注这些Twitter吧!

我们是谁,我们是如何涉足数字化历史领域的?

同许多历史学家相似,伊恩·米利根是以迂回的方式进入数字化历史领域的。他小时候是一名电脑极客:用BASIC(一种编程语言)编程、设计电脑游戏、在车库里销售硬件设备。在这一点上,他是人数众多的"中产阶级白人"中的一员,这些人从小就有机会接触到计算机,同时也会思考编码是否是探究数字人文必不可少的一部分,并对此有着清晰的认识,正如米丽亚姆·波斯纳(Miriam Posner)曾好心提醒过的那样。[①] 一两个与高等微积分相伴的灾难性学期和与加拿大军队渐行渐远的经历使米利根远离了与计算机有关的工作, xxii
对历史着了迷,并在几年后攻读博士学位的后期重新察觉了自己钟爱的事情。

在攻读博士学位期间,米利根研究了青年文化,其中少不了对大量的个体进行研究。当他在推进第二个项目却不知道下一步该怎么做时,一次偶然的机会,他与数字化历史学家威廉·特克尔(William Turkel)进行了交流,并参

① 米丽亚姆·波斯纳曾表示:"在劝说所有人研究代码之前需要思考一些问题。"

加了一场人文科技联合会议——THATCamp，这使得他最终走上了数字人文这条路。他首先借助《编程历史学家》这本书学习了文本分析和数据挖掘技巧，然后经历了数月的试验和试错。这一切都意义重大，因为如今他正在滑铁卢大学担任数字化历史的助理教授，那是一所专注于科学和工程的学校。他给学生的建议是：只要勇于面对失败，任何人都可以从事计算工作。

肖恩拥有与伊恩极为相似的背景：他的第一台电脑是他哥哥攒钱买的Com-modore Vic-20。当时游戏对他们来说是稀罕玩意儿，所以他和哥哥开发了一个工作流技术，用来读取《电脑》(*Computer*)等杂志上发布的代码，输入代码并进行错误检查。如果他们足够幸运，两周后就会拥有游戏《太空入侵者》(*Spaceinvaders*)的克隆版本。大学期间，肖恩成功意识到“做学问”唯一的含义就是写论文。1995 年他被要求利用新兴的“万维网”来获取有关伊特鲁里亚人的信息，肖恩对此非常沮丧但仍尽职尽责地写了一篇论文，他在第一行就写道：“万维网永远不会对搞学术有帮助。”

几年后他就食言了，肖恩的博士论文与罗马制砖业的碑文和考古学相关，该研究迫切需要使用数字化工具进行网络分析，因为在当时这是唯一有助于理清复杂关系的方法。肖恩重新找回了儿时开发电子游戏的动力并努力使这些网络重焕生机(阅读了有关人工智能的材料，肖恩发现开发基于主体的偶然性模型极具潜力)。肖恩在马尼托巴大学获得了博士后研究员职位，那是利·斯特林(Lea Stirling)就这个奇怪想法上进行冒险尝试的地方。这项工作在林肯市内布拉斯加州大学举行的第一次“数字人文工作坊”上得到了认可，当时艾伦·刘(Alan Liu)和威·G. 托马斯三世(William G. Thomas Ⅲ)建议他应该在博客上介绍他的工作，http://electricarchaeology.ca/应运而生。肖恩利用他的博客探索用于考古学教学和研究的新型工具(记录下了他成功和失败
xxiii 的经历)，这成为他开展研究以及在网络教育领域的基石。在取得博士学位八年后，他在卡尔顿大学历史系获得了数字人文学助理教授的职位。

斯科特打小就对历史充满热情，同时致力于计算机事业，他相信历史和计算机终将成为自己职业生涯不可或缺的元素。斯科特花了三年半时间攻读计算机工程学位，但在这期间他对工程的失望却与日俱增，并发誓再也不会涉足计算机领域。他研修了与工程学课程数量相当的历史学课程，并较为轻松地取得了科学史专业的学位，他很乐意在余生的大部分时间里与尘土覆盖的旧

档案打交道。当他申请印第安纳大学并同时获得攻读科学史和信息科学研究生学位资格之时，斯科特意识到信息科学会有助于历史学研究，所以他开始攻读双学位。在很短的时间内，他了解到了数字人文的存在，并认为将计算机应用于自身感兴趣的地方十分有趣。

作为一名研究科学史的学者，斯科特对建模和绘制科学的发展产生了兴趣。他对科学曲折发展以及所有研究都有所涉猎却停滞不前的时段感兴趣。这些领域往往被非常宽泛的术语概括且缺乏相应的深入研究。斯科特的研究希望将宏观视野同微观数据结合起来，从而同时洞察宏观和微观。由于这种计算方法十分罕见，所以他开始开发相关的软件技术和软件包以帮助他实现目标。他的导师极有远见，特别是历史学家罗伯特·A. 哈奇（Robert A. Hatch）和信息科学家凯蒂·博内尔（Katy Börner），他们在帮助他实现这些目标的过程中发挥了重要作用。

斯科特将探究历史、数据和数字人文的历程不定期地记录在他的博客上。他目前在卡耐基梅隆大学担任数字人文专家，同时还在印第安纳大学攻读博士学位。

将我们的个人经历联系在一起就会发现一个关键的共同点，即尽管走了一些弯路，但我们都渴望透过现象看到本质。如果你也如此，那么让我们从第一章开始数字化历史学历程。

目　录

Exploring Big Historical Data

The Historian's Macroscope

“如果你对于开启一个数字人文项目的方式和原因很好奇，这本书可以解答你大部分的问题。本书通俗易懂，简洁明了，有利于你探究数字人文，同时为数字人文领域之外的方法论问题带来了意义深远且发人深思的启示。”

美国西点军校历史学家 丹尼尔·弗兰卡博士

“毫无疑问，我会在数字化历史学的课堂上使用本书，尤其在我介绍网络可视化之时。本书很好地定义与阐述了数字人文领域的专业语言，拉近了其与学生和历史学家的距离。在我所阅读过的数字人文资料中，本书对于网络分析基础知识的讲解是最清晰的。”

美国加州大学洛杉矶分校数字人文部核心教师 贾尼斯·L. 赖夫教授

“互联网是一种物理网络结构这一理念，以及与之相关的网络和网络交互理论，对于很多本科生来说极其新颖，对于一些人而言学习起来具有相当的挑战性。一直以来，学界缺乏一本对于大学生来说相对友好的网络分析介绍。本书填补了这一空白。”

英国伦敦大学国王学院 斯图尔特·邓恩博士

ISBN 978-7-308-18858-6

9 787308 188586

定价：59.00元