

北大高等教育文库
• 学术规范与研究方法丛书 •

INTERNET RESEARCH SKILLS



[爱尔兰] 尼奥·欧·杜恰泰 (Niall Ó Dochartaigh) 著

如何利用互联网做研究



北京大学出版社
PEKING UNIVERSITY PRESS

如何利用互联网做研究

[爱尔兰] 尼奥·欧·杜恰泰 著
傅存良 译



北京大学出版社
PEKING UNIVERSITY PRESS

著作权合同登记号 图字:01-2014-4165

图书在版编目(CIP)数据

如何利用互联网做研究/(爱尔兰)杜恰泰著;傅存良译. —北京:北京大学出版社,2014.10

(北大高等教育文库·学术规范与研究方法丛书)

ISBN 978-7-301-24895-9

I. ①如… II. ①杜… ②傅… III. ①互连网络—应用—社会科学—研究方法—教材 IV. ①C39

中国版本图书馆CIP数据核字(2014)第228763号

Internet Research Skills.

English language edition published by SAGE Publications of London,
Thousand Oaks, New Delhi and Singapore.

© 2012 Niall Ó Dochartaigh

© 2012 Chapter 7 Niall Ó Dochartaigh and Patricia Sleeman

书名:如何利用互联网做研究

著作责任者:[爱尔兰]尼奥·欧·杜恰泰 著 傅存良 译

丛书策划:周雁翎

丛书主持:泮颖雯

责任编辑:刘军

标准书号:ISBN 978-7-301-24895-9/C·1056

出版发行:北京大学出版社

地址:北京市海淀区成府路205号 100871

网站: <http://www.pup.cn> 新浪官方微博: @北京大学出版社

电子信箱: zpup@pup.cn

电话: 邮购部 62752015 发行部 62750672 编辑部 62767346

出版部 62754962

印刷者:三河市北燕印装有限公司

650毫米×980毫米 16开本 14.25印张 210千字

2014年10月第1版 2014年10月第1次印刷

定价:38.00元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子信箱: fd@pup.pku.edu.cn

目 录

第一章 导论	1
在云里:信息时代研究新攻略	1
搜索引擎的兴起	3
各章概要	6
本书目标读者	8
第二章 著作	10
网上印刷出版物	10
图书馆目录	17
网上书店	21
网上全本图书	23
书评	31
出版社	32

第三章 文章	36
认识学术文章	36
评论年鉴	40
谷歌学术搜索	40
索引和摘要数据库	44
专业学科数据库	54
网上文章全文	58
出版社	61
开放存取	63
其他文章数据库	66
硕士、博士论文	67
第四章 主题指南	73
检索开放网络	73
学术主题指南	76
通用主题指南	81
认识网址	85
认识网站	89
第五章 检索关键词搜索引擎	95
导论	95
三巨头	97
高级检索	100
元搜索引擎	104
搜索引擎与元搜索引擎对比	105

布尔检索	105
搜索引擎如何检索	106
设计检索查询	108
三思而行	111
技术难题	112
第六章 社交媒体、新闻和多媒体	117
社交媒体	117
电子邮件列表	126
讨论组	129
新闻	131
多媒体	135
第七章 政府、档案和统计资料	145
导论	145
政府	146
档案	153
统计资料	159
第八章 评价和引用	168
导论	168
分类	169
评估	174
引文	178

第九章 超越检索:网络、研究管理和在线研究	191
研究网络和研究支持	191
社交网络	192
整理研究	196
互联网研究方法	208
附录 社会科学和人文学科的学术博客	213
参考文献	217

第一章 导论

在云里:信息时代研究新攻略

数年前诞生的“云计算”一词,指的是在由众多计算机互联而成的在线“云”里,进行数据和文件的存储加工工作。“云”也是对互联网以及当下我们所处的新信息环境的绝妙比喻。我们被信息包围,饱受信息轰炸。在某种程度上,我们受到信息的遮蔽。海量信息会闭目塞聪、妨碍理解。随着信息量增加,辨识什么信息靠得住、什么信息算创见,尤其是什么信息重要——这些工作越来越有挑战性。很多人认为,新信息技术洪流所产生的问题,对研究学问、追求学识提出了全新考验。《谷歌让我们变蠢了?》一文——后来扩充为专著,名为《浅薄》(Carr,2010)——认为,互联网的应用,正在改变我们接收信息的方式,因此也在改变我们的思维方式。作者指出,技术允许我们在零碎信息间快速游走,这种方式使我们丧失了接受和理解长篇复杂信息所需的技能。近来众多著作都提出了类似观点,比如《走神》(Jackson,2009)、《大脑超载》(Klingberg,2009)、《数字化一代》(Tapscott,2009)等,也许最刺激的当属《极蠢一代》了,其副题为

“数字时代麻醉美国年轻人、危及我们的未来”(Bauerlein, 2008)。这些著作认为,我们处在信息接收方式和思维方式的深刻历史变迁之中。尽管恐慌感、危机感以及互联网末日感也许有些言过其实,但研究者显然面临着全新挑战。在这种新环境下,最重要的研究方法在于信息的沟通、评估、遴选和限定,而非简单收集大量相关信息。

近现代社会科学和人文学科的奠基者们于18世纪末至20世纪初著书立说。在他们所处的时代,学者差不多还能做到在相对宽泛的领域或论题内博闻强识,尚有望就其学科“悉数”通读。他们还如饥似渴地从其他领域寻求新观点。那些早期作家认为,包罗万象的阅读和研究方法非常重要,西格蒙德·弗洛伊德初版于1900年的奠基之作《释梦》即为明证。该书开头就是一段简洁的文献概述,弗洛伊德将解梦释梦法一直追溯到亚里士多德,追溯到古希腊原文中提及梦的最早书面材料,其概述同样跨越了广博的学科。弗洛伊德强调,尽管那样,现有论述梦的著作还是过于庞杂,令他在概述中难以无所不包,而其一网打尽的雄心也显露无遗。他想确保没遗漏任何重要资料,没错过一篇相关文献,无论源自哪门学科。

到了20世纪中叶,学术研究和学术刊物爆发式发展,产生了卷帙浩繁的研究,致使任何人都别想针对某一特定学问的主要分支通览群书。20世纪70年代,“信息过载”一词开始家喻户晓,而互联网还远未出现。过去二十多年来,事关新技术的多种因素导致研究刊物指数级增长。20世纪早期,研究者的核心任务是寻找、检索与论题有关的各种潜在重要资料和专著;而到了21世纪,研究者的核心任务则是有效地限制、控制和引导种种信息流。

这里我们可以拿其他类型的资源打个实用比方以便于理解。近几十年来,在选择方面以及随之而来的强调质量评价、限制和控制的重要性方面,信息领域并非唯一出现指数级增长的领域。数千年来,人类最重要的斗争就是猎杀、种植或收集足够的食物以延续生命;过去半个世纪,发达国家食物过剩、取之不

尽,彻底改变了人类与食物关系的某些基本要素。中世纪国王身材滚圆,其肥胖身躯象征着繁荣富足;如今取而代之的修长健壮的形象,则是对节食习惯和自制力的颂扬。如今,发达国家有关营养学最重要的核心法则旨在约束食物摄入,并在摄入食物时更为自觉。与此类似,研究者发现,为适应看上去无穷无尽的“信息云”,对阅读及信息摄入的限制、限定、约束必须比以往严之又严。就某一课题“有所”发现的能力,或就某一课题发现海量材料的能力,其重要性在减弱,而对这些海量材料进行筛选的能力以及识别少量关键资料和文本的能力成了核心技能。如今,研究者必须具备比以往更强的信息检索技能,并且更需要具有学科领域的专门知识,以便识别研究领域内关键的核心资源并加以突出。同时,研究者也更加需要和其他同行建立网络,交流经验,分享知识。

搜索引擎的兴起

大型关键词搜索引擎,如“谷歌”(Google)及其学术分支“谷歌学术搜索”(Google Scholar),如今已成为核心研究手段,而仿照传统图书分类法查找资料的替代做法证明是行不通的。图书馆和研究院力图遏制谷歌成为学术研究工具,并通过按学科领域对网上学术资源进行系统化的编目、整理来推广替代检索工具,这一努力历时 15 年,即将宣告失败。这个庞大的离心方案,由各种不同的项目和机构于 20 世纪 90 年代开始实施,试图复制图书馆按学科和子主题进行图书编目、整理的职责,以帮助学生和研究者找到他们所需的材料。这个方案如今几乎完全终止了。2011 年出现了里程碑式的重要事件:受到悉心维护的英国大型学术主题目录网站 Intute 最终希望破灭,财源耗尽,关门大吉(第四章)。但是,对大型搜索引擎与日俱增的依赖也有自身的问题。近年来,互联网在线研究的方法漫无目的,在很大程度上导致学生使用劣质和不可靠的资料的现象激增。2004 年,2300 多位美国大学教授接受民意调查,超过 42% 的人表示,

学生作业的质量因为用了互联网而变差,只有 22% 的人认为质量提高了(Jones and Johnson-Yale, 2005)。不过,大型搜索引擎仍会处于研究过程的中心,而图书馆不会重新成为绝大多数学生和学术研究人员的主要资料来源了。

小心翼翼、有条不紊的编目工作受到双向夹击。一方面,关键词搜索引擎(其中主要是谷歌)日趋完善,甚至打字时就在猜我们想要查什么,根据我们以往的检索模式,识别我们可能感兴趣的资源,把最重要的资源在查询列表中置顶。如果你对某一特定学科领域的学术研究资料有固定的检索模式,搜索引擎就会相应调整你的查询,将此类资源在查询结果中置顶。如今,关键词搜索引擎可以更快地列出大量同类站点,再去浪费时间浏览一大串网站就越发没理由了。

按关键词检索胜过大型编目工程,但是按主题检索并未消亡。它只是呈现出新形态:更专业、更有个性,也更分散。一方面,大型编目受到关键词搜索引擎的伏击,另一方面,它们又遭遇了按学科领域整理材料的新方法,这些新方法产生了更有重点、更为专业的主题目录,而不必使用“编目”这个术语。例如,大学授课课程和阅读书目,就在非常精密的子领域内起到了文献指南的作用。如今,通过一系列公开课件方案(第二章),它们比以往得到更普遍的利用。每篇学术文章结尾的参考书目相当于另一种关于某主题读物的专业指南。我们现在享用的学术数据库之间的无缝链接,使得从兴趣领域内最重要的文章出发去查阅相关读物变得轻而易举。比起在同一主题栏下查找分类材料(第三章)来,这种主题检索方法更简单,通常也更有效。合理限定阅读材料最有效的策略之一,就是只要沿着研究领域内的参考书目路径进行即可。先听取周围或在线联系的同事、导师、教师以及研究人员的意见,确认论题内的几本关键书刊,再及时翻来覆去沿着参考书目的路径,从参考书目中寻找那些关键著作,看看那些重要作者引用了哪些著作,然后搜寻与研究主题直接相关的内容。及时前进,检索引用过与课题联系最紧密的文本的文章,用一两个与主题相关的词限定检索,这样有助

于略去不相干的结果。你几乎可以在包括学术研究刊物的任何一个网络服务平台上进行这两种检索。接下来的两章会详细论述这些服务。然后你可以按照参考书目前前后后寻找刚开始检索时出现的最重要的著作和文章。在进行类似检索的过程中,你开始划出了自己特定的工作领域,并明确了研究项目的边界。

社交媒介和“用户原创内容”(user-generated content)也导致各种新型专业化指南激增。其资源与学术博客、亚马逊(Amazon)的“迷书单”(Listmania)(第二章)、“脸书”(Facebook)以及“学术网”(academia.edu)之类的学术联网网站(第九章)同样丰富,通过联合网络达人的成果,兼收并蓄相关资源。随着大型主题目录的边缘化,关键词搜索引擎和界定非常鲜明的学科领域内的专业化资源指南,可按多种多样的形式加以应用,如今已成为网上检索和研究的两大支柱。

认识互联网

互联网并非有组织的体系。没人在看管互联网。它原本并非网络,更非众网络之网。首先,互联网其实很简单,就是全世界成千上万的电脑可以彼此联通。你在自己的电脑屏幕上点击一个链接,你的电脑就向另一台电脑发出一条信息,让它把链接所代表的文件传给你。另一台电脑会理解这个指令并发送文件。本质上,互联网由其他人电脑上的文件组成,包括文本文件、音频视频文件,这些文件允许外部世界查看。把信息放到“互联网上”,人们只需将文件移动到其电脑的某部分或别的电脑上,也就向外部世界公开了。

输送指令的网络,包括电话线、卫星天线和电缆。此类通信线路早就有了。其中大多数并非专为传输互联网信息而建,它们也并非互联网的关键。

20世纪60年代开始有了互联网。也就是说,从那

时起,电脑就能长距离传递信息了。1993年,万维网(World Wide Web)诞生,这让更多公众更方便地由此获益。网络基于http,即“超文本传输协议”(the Hyper Text Transfer Protocol),它推出的最主要创新是超文本,即能让你连接到其他网页的网页链接。嵌入每个超链接的是互联网地址,即URL。它能显示与你相连的计算机上的文档或图像,也可以显示世界另一端计算机上的文档或图像。它让人们头一回在网上制作极为复杂的文档,包括文档、图表、声音以及视频的链接。

网络与人们在其电脑上公开的信息不能混为一谈。人人都会提供各自的信息。网络“协议”仅仅表示所有机器都彼此理解,并以标准方式展示来自其他机器的文档。苹果浏览器(Safari)、火狐(Firefox)或微软浏览器(Internet Explorer)之类的网络浏览器对上述信息并不负责——它们不过是提供了一个查看信息的捷径。

各章概要

本书先用两章集中讨论网上学术著作和文章,将用户引向最有序、最可靠的现有资源,接着在随后的章节继续论述大型数据库之外的开放式网络在线资源,这些资源远远谈不上条理清晰,更难有效检索和评估。针对一系列研究任务,每章都结合多种网上资源,用实例说明检索方案。通过文献检索,这些方案使读者从最初对其论题的探究,进而能鉴定重要资料,达到专业化检索查询和查找原始数据源及文献资料的新阶段。这些策略带来了各种各样的网上资源,这强有力地证明,许多人完全依赖的大型搜索引擎,仅能提供访问一小部分网上可用研究资源。

第二章集中讨论网上学术著作。一系列网上服务,包括亚马逊之类的书店和“谷歌图书搜索”(Google Book Search)之类

的检索服务,提供了获取全本学术著作的新路径,而电子书允许研究者以新方式注释、整理、检索学术著作。该章概述了结合不同网上资源,利用网上图书馆目录、出版社网址和网上教材进行图书检索的策略,这在几年前是无法做到的;还概述了“大海捞针”式的检索法,这种检索法能指导你找到那些光看标题和卷目不会留意的书籍,指导你如何利用那些服务辨别研究领域内的重要著作,如何将诸如读者评论等新资源纳入检索策略。

第三章主要涉及学术文章。该章阐明了如何综合运用摘要数据库、全文数据库、图书目录、“谷歌学术搜索”以及类似工具充分利用现有资源;阐明了这些服务的一致部分及其方法上的差异,对比了它们之间的优劣,理清了竞争伙伴的重叠内容;评估了致力于特定学科领域的商业数据库,并强调了论文的价值;还论述了越来越多获取期刊论文、硕士论文、博士论文、工作报告、会议论文及其他学术材料的方法,多亏了“开放存取运动”(the open access movement)和公共储存库的壮大,这些方法才得以奏效。

第四章是论述“开放式网络”的头一章。和学术著作、学术文章数据库相比,“开放式网络”是个混乱的领域。该章概述了如何应对资源权威性和可信度不明朗的信息环境,以及在条理、结构、界限上不如期刊文章和著作数据库清晰的资源,讨论了按学科领域整理材料的网上学术资源指南,并概述了在研究领域内辨识网上关键资源的策略,描述了网站是如何运作的,详尽解释了能让读者加深认识所查材料及其背景的技术。

谷歌、雅虎(Yahoo)和必应(Bing)等大型关键词搜索引擎在第五章加以论述。该章讨论了为个体研究者查询结果强化定制服务的意义,讨论了高级检索选项,就分析结果提出建议,解释了搜索引擎的搜索方式和搜索内容,就编制查询提出建议,并解释了如何有效高效利用这些搜索引擎。

术语“网络 2.0”(Web 2.0)和“网络 3.0”(Web 3.0)描述的是以其互动性为部分特征的新一代网上服务和网上资源。这些服务大多模糊了作者与读者、制作者与消费者的区分,把观众

和读者的贡献纳入网上服务,制造强大新颖的信息分类、定位方法。第六章关于“社交媒介、新媒体和多媒体”,通过新闻资料和一系列现有网络视频音频资源,对这些服务进行了论述。

第七章关注政府资料、档案和统计资料。也许新技术最引人瞩目的作用就是打开了获取这些原始资料的通道,能让我们查询数以百万页的立法辩论,能让我们获取统计报告,在很多情况下还能获取统计报告所依据的原始数据,让档案变得更加透明。该章鉴别了查找这些原始资料的重要资源,并提出了研究策略。

第八章着手列举评价和认识网上出处的指导原则,提供了一系列问题,适用于查询能在网上找到的任何材料,并就引用各类网上材料提供了详细的建议。

本书最后一章就整理研究、管理研究笔记和读物提出建议。该章比较了各种资料管理程序包,衡量了把全部研究资料装进一个软件包的利弊,论述了将研究生和研究者联系在一起的在线网络,讨论了专门致力于学术联网的最新产品和服务,最后简要概述了网上研究方法——当前把互联网本身当做社会变革和政治运动场所的研究者运用了其中的许多方法。该章还探讨了由网上检索和网上互动产生的有关社会变革的新型资料,探讨了研究者如何开始运用这种资料处理社会属性和政治属性方面的新旧问题。

本书重点研究的是英语资源,旁及英语在内的多语资源。每种语言都会另有一系列重要资源,这些资源仅对查找该语种资源的人重要,不过这些资源并未包括在本书范围之内。

本书目标读者

本书可以作为本科生及研究生互联网研究攻略的课程教材,这些课程带领学生进入人文学科和社会科学,把互联网当做研究资源加以高效率、批判性的利用。对于关注当前辩论和论战的议题,研究媒体、政府乃至志愿组织等广泛领域的人而言,

本书也不无裨益。本书假定读者经常利用互联网进行研究,不过有时候会一无所获、垂头丧气。

对于第一次撰写拓展论文或从事独立研究任务的高年级本科生来说,本书益处良多,尤其是他们希望继续研究生学业的话。本书特别为着手撰写硕士论文和博士论文的研究生量身定制。对于博士论文头一年撰写文献综述、探索研究领域、识别重要资源的学生来说,本书实为良伴。本书专为包括撰写硕士、博士论文及相关训练在内的一系列工作提供了建议。

本书对经验丰富的研究者也有价值,他们虽然经常运用互联网,但从未系统地探索其领域内的可用资源,从未形成进行网上研究的系统策略。

每章结尾的练习题带领读者重温正文涉及的内容,让他们能在自己的研究领域或特殊兴趣领域内自由钻研。通过练习,读者可以不断完成与他们的研究项目有关的众多重要工作。全书强调了批判看待资料来源的重要性,强调了对网上资料来源必须形成全面认识的重要性。

本书能帮助读者安排好自己的网上研究,辨别其领域内的重要资源,研究具体任务时灵活综合各类网上资源,更有效、高效地运用大型搜索引擎,批判性评估网上资源,在研究中更充分地利用网上资源的巨大潜能。

第二章 著作

网上印刷出版物

谈论互联网和铅字,有时就像是在谈论两种截然不同的信息,其中一个不可靠、令人生疑,另一个靠得住、值得信赖。近些年,这一区别被彻底打破,把互联网和铅字当成两种对立范畴也不再有任何意义。随着技术与经济的飞速变化,互联网如今已是发布、获取学术刊物文章的最重要媒介,并迅速成为发布著作、获取著作的主要媒介。如果你在寻找印刷出版物,最好就从网上开始。

尽管网上之物与印刷之物的区别被打破了,印刷出版物的概念仍然很重要。网上文献也有印刷版,印刷版通常要经受印刷出版领域形成的一套质量把关措施的检验,而纯粹网上文献大多用不着如此检验。印刷出版物更有可能要走同行评审程序,业内专家由此评估是否值得出版;要走编辑程序,作者会被劝说改进作品;要通过审稿和校对程序,以纠错勘误。这些程序远非无懈可击,也遭受许多尖锐、合理的批评(Harley and Acord, 2011)。即便同行评审过了关,印刷出版物也会像网页一样文字