

在线信息搜寻行为的 非介入性研究

姜婷婷 著



科学出版社

在线信息搜寻行为的 非介入性研究

姜婷婷 著

科学出版社

北 京

内 容 简 介

本书围绕在线信息搜寻行为展开,分为理论篇、方法篇和实证篇三个部分。

理论篇阐述了信息搜寻行为研究的背景、意义和发展历程,系统梳理了该研究领域内的关键概念和经典理论模型,为实证研究奠定了理论基础。

方法篇是本书的重点,从通用研究方式和方法论入手,深入分析了各种研究方法在信息搜寻行为研究中的应用,特别采取介入性特征对诸多方法进行分类,从而引出本书所关注的核心——事务日志分析这一非介入性方法。该方法属于网络分析的范畴,进一步可以划分为搜索日志分析和点击流数据分析。基于对网络分析的总体介绍,本书着重展示了搜索日志分析和点击流数据分析这两种方法研究框架的构建及其具体分析过程与指标。

实证篇是对非介入性的事务日志分析方法的实际应用,分别基于传统的学术图书馆 OPAC 系统和新兴的社会性图书馆系统这两种信息环境,针对其用户的信息搜寻行为进行了日志数据的采集、处理和分析,使得以上研究框架的有效性和适用性得以验证,为信息搜寻行为领域贡献了一系列通过其他方法难以获得的研究发现,同时也为网络信息环境的设计提供了重要启示。

本书适合从事信息管理、情报学、图书馆学、信息构建和用户体验设计等领域的理论和实践工作者阅读参考,也可以作为高等院校相关专业硕士和博士研究生的教学参考书。

图书在版编目 (CIP) 数据

在线信息搜寻行为的非介入性研究 / 姜婷婷著. —北京: 科学出版社, 2017.9

ISBN 978-7-03-054622-7

I. ①在… II. ①姜… III. ①信息检索—研究 IV. ①G254.9

中国版本图书馆 CIP 数据核字 (2017) 第 239232 号

责任编辑: 余 丁 阆 瑞 / 责任校对: 郭瑞芝

责任印制: 张克忠 / 封面设计: 迷底书装

科学出版社出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

中国科学院印刷厂印刷

科学出版社发行 各地新华书店经销

*

2017 年 9 月第 一 版 开本: 720×1000 B5

2018 年 1 月第一次印刷 印张: 17 3/4 插页: 6

字数: 370 000

定价: 98.00 元

(如有印装质量问题, 我社负责调换)

前 言

信息搜寻行为作为情报学的重要研究分支,旨在探索人类获取信息的外在行为表现和内在影响机制。以往的信息搜寻行为研究过多地依赖问卷、访谈、实验等方法,这些方法都是典型的介入性方法,研究对象可能因为意识到自己正置身于研究之中而做出不真实的行为或是对自己的行为做出不准确的表述,从而降低研究的效度。正是考虑到了这种反应性问题的存在,本书首次立足于信息搜寻行为领域探讨如何引入非介入性研究。

非介入性研究在社会科学中由来已久,最早于1966年由Webb等在《非介入性测量:社会科学中的非反应性研究》一书中提出。非介入指的是在研究的数据采集阶段,被研究的现象未受到研究本身的干扰。非介入性研究的核心思想是利用过去行为留下的痕迹数据,既包括语言类数据,又包括行为类数据,这些数据都是人们在现实世界中自然从事各种活动时留下的,而不是为了特定研究目的而产生的。利用痕迹数据开展研究能够避免或减少外界的影响,从而弱化观察者效应、消除观察者偏见。随着互联网逐步渗透到人类生活的方方面面,数字化的痕迹数据也越来越丰富,例如,人们主动分享的照片、视频资源或在论坛、博客、微博上留下的文字信息等。除了这些可以直接在网页上抓取的数据,网站服务器上的事务日志在后台记录了用户所有的使用行为,是研究在线信息搜寻行为的理想数据来源。

本书的研究重点正是事务日志分析在信息搜寻行为研究中的应用。搜索日志分析和点击流数据分析都是事务日志分析的重要形式。前者在网络搜索行为研究中比较常见,由Jansen提出的搜索日志分析过程已经获得了广泛认可,相关研究一般遵循数据采集、处理和分析三个阶段,并在最后一个阶段分别针对关键词、查询式和搜索会话开展分析。点击流数据分析则多见于电子商务领域,重点关注顾客的商品浏览和购买行为特征。然而,由于缺乏统一的分析框架,该方法一直未能在其他领域获得广泛应用。实际上,在线信息搜寻行为研究与在线购物行为研究具有相似的模式:顾客访问在线购物网站的目的是购买商品,通过研究他们浏览了哪些商品、是否点击了广告、是否付款购买等情况,可以了解网站的销售模式和营销策略是否有效、哪些功能是有利可图的以及界面设计中存在哪些问题等;信息搜寻者访问各类信息富集型网站的目的是获取信息资源,通过研究他们对哪些信息资源感兴趣、如何查找资源、是否找到了资源等情况,可以揭示网站信息构建的有效性及其提供的信息搜寻体验。

因此,本书全面调查了各种背景下的点击流数据分析研究现状,从中抽取出常

见的分析思路和基本指标,此外还特别关注了点击流数据的可视化分析手段。在此基础上,本书指出点击流数据分析过程也应该由采集、处理和分析三个阶段组成,并针对数据分析创建了一套全新的通用分析框架。该框架分为足迹、移动和路径三个层次:每个被请求的页面都因为用户的到访而产生了足迹;用户的每一次点击都将他从当前页面带到了链接所指向的页面,使他在网站中发生了移动;将用户请求的所有页面按照时间顺序连接起来就得到了他在网站中导航的路径。每个层次都包含了一系列分析维度和指标,基于不同的粒度反映用户信息搜寻行为。丰富的可视化分析手段也是该框架的一大特点,尤其是在路径层次上开创性地引入了桑基图展现用户群体路径,并创建了路径图可视化展现用户个体路径。本书将该框架分别应用于以学术图书馆 OPAC 系统和社会性图书馆系统为背景的实证研究中,其适用性和有效性都得到了验证。

全书共 10 章。第 1 章“导论”介绍了网络信息环境的演进对人类信息搜寻活动带来的影响和信息搜寻行为研究所经历的四个发展阶段,提出了非介入性方法对于本领域研究的重要价值。第 2 章“信息搜寻行为相关概念”和第 3 章“信息搜寻行为相关模型”为本书的理论基础部分,对信息、信息需求、信息搜寻等概念进行了阐释与辨析,也对领域内的高被引理论模型进行了描述和应用分析。第 4 章“研究方法论”和第 5 章“信息搜寻行为研究方法”共同组成了本书的方法基础,基于不同的研究方式系统介绍了通用方法论,再根据方法介入性对信息搜寻行为领域常见研究方法进行分类,详细介绍了各方法的特点与实际应用。第 6 章“事务日志分析基础:网络分析”、第 7 章“搜索日志分析方法”及第 8 章“点击流数据分析方法”是本书的核心章节,展现了基于页面标签的网络分析以及基于服务器日志的搜索日志分析和点击流数据分析这三种研究方法的全貌,从研究过程到分析层次与具体指标,为在线信息搜寻行为研究提供了非介入性的方法论体系。最后,第 9 章“学术图书馆 OPAC 系统用户信息搜寻行为研究”和第 10 章“社会性图书馆系统用户信息搜寻行为研究”分别选择了 Web 1.0 和 Web 2.0 时代具有代表性的网络信息环境作为研究背景,利用事务日志分析方法对其用户信息搜寻行为开展非介入性研究,同时也通过实证研究对前面章节中提出的数据分析框架进行检验。

本书是在作者的博士毕业论文 *Characterizing and Evaluating Users' Information Seeking Behavior in Social Tagging Systems* 基础之上,补充了近几年的最新研究成果撰写而成。在此,特别表达对博士导师——美国匹兹堡大学信息科学学院 Sherry Koshman 教授的真挚谢意和无限怀念,是她帮助作者在信息搜寻行为领域开创了一条系统的研究路径,虽然她已经离开了人世,但是她的学术热忱与风范将永远影响作者。

本书是国家自然科学基金项目“用户探寻式搜索策略分析及系统构建研究”(项目编号:71203163)、教育部人文社会科学研究项目“社会性标签系统中用户信息搜

寻行为研究”(项目编号: 12YJC870011)和国家自然科学基金重点国际(地区)合作研究项目“大数据环境下的知识组织与服务创新研究”(项目编号: 71420107026)的成果之一。

作者的硕士研究生迟宇、史敏珊、范水香、王淼、陈舜昌、贺虹虹、杨雨琪、徐亚苹、郭倩和杨佳琪参与了以上项目的研究,在实证研究的数据采集与分析方面做了大量工作,并发表了与项目相关的学术论文,为本书的撰写奠定了重要基础。在本书初稿完成后,徐亚苹、郭倩和杨佳琪承担了整理和编排工作,在此表示衷心感谢。

尽管作者在项目研究和本书撰写过程中付出了艰苦的努力,但不足之处在所难免,恳请读者不吝赐教,以使本书更臻完善。

姜婷婷

2017年7月于珞珈山

目 录

第 1 章 导论	1
1.1 在线信息搜寻活动	1
1.2 信息搜寻行为研究的发展历程	2
1.2.1 萌芽阶段 (20 世纪 40、50 年代)	2
1.2.2 起步阶段 (20 世纪 60、70 年代)	3
1.2.3 形成性发展阶段 (20 世纪 80、90 年代)	4
1.2.4 爆炸性发展阶段 (2000 年以后)	5
1.3 非介入性研究的兴起	7
1.3.1 非介入性研究的起源与发展	7
1.3.2 引入非介入性研究方法的意义	8
1.4 本书的内容与意义	10
1.4.1 内容概述	10
1.4.2 研究意义	12
第 2 章 信息搜寻行为相关概念	14
2.1 信息	14
2.1.1 “外在-内在-意义建构”信息分类法	14
2.1.2 “过程-知识-事物”信息分类法	15
2.1.3 “环境-个人-社会”信息分类法	17
2.1.4 “商品-数据-知识-过程”信息分类法	18
2.2 信息需求	19
2.3 信息搜寻	21
2.4 信息搜索	23
2.5 信息偶遇	23
2.6 信息使用	24
2.7 信息规避	25
2.8 信息素养	25
2.9 信息实践	26
2.10 本章小结	27

第 3 章 信息搜寻行为相关模型	28
3.1 方法论模型	28
3.2 Krikelas 的信息搜寻行为模型	31
3.3 Bates 的信息搜寻模式模型	33
3.4 Ellis 的信息搜寻行为模式模型	35
3.5 Kuhlthau 的信息搜索过程模型	37
3.6 Marchionini 的信息搜寻过程模型	39
3.7 Wilson 的信息行为模型	42
3.8 Byström 和 Järvelin 的信息搜寻模型	44
3.9 Johnson 的信息搜寻综合模型	46
3.10 Savolainen 的日常生活信息搜寻模型	48
3.11 Williamson 的日常生活信息获取生态模型	51
3.12 本章小结	52
第 4 章 研究方法论	53
4.1 定性研究方法论	54
4.1.1 个案研究	54
4.1.2 人种学研究	55
4.1.3 扎根理论研究	56
4.1.4 叙事研究	58
4.1.5 现象学研究	59
4.2 定量研究方法论	60
4.2.1 调查研究	60
4.2.2 实验研究	61
4.3 混合方法研究模型	63
4.4 本章小结	64
第 5 章 信息搜寻行为研究方法	65
5.1 介入性研究方法	66
5.1.1 问卷	66
5.1.2 访谈	69
5.1.3 焦点小组	71
5.1.4 日记	73
5.1.5 实验	75
5.2 非介入性研究方法	76
5.2.1 内容分析	76

5.2.2	元分析	78
5.2.3	元人种志方法	80
5.2.4	观察	82
5.3	本章小结	84
第 6 章	事务日志分析基础：网络分析	85
6.1	网络分析概述	85
6.2	网络分析的数据采集策略	86
6.2.1	服务器日志	87
6.2.2	页面标签	88
6.2.3	服务器日志与页面标签方法比较	90
6.2.4	其他数据采集方法	91
6.3	网络分析的数据分析策略	92
6.3.1	基本量度指标分析	92
6.3.2	Google Analytics 分析维度	94
6.4	网络分析的辅助方法	100
6.4.1	可用性测试	100
6.4.2	启发式评价	101
6.4.3	问卷调查	103
6.5	本章小结	104
第 7 章	搜索日志分析方法	105
7.1	搜索日志分析方法探索	105
7.2	搜索日志分析过程	107
7.2.1	搜索日志数据采集	107
7.2.2	搜索日志数据准备	108
7.2.3	搜索日志数据分析	110
7.3	搜索日志分析研究实例	111
7.3.1	元搜索引擎 Dogpile 日志分析	111
7.3.2	中文搜索引擎 Timway 日志分析	117
7.4	搜索日志分析研究现状	124
7.4.1	通用网络搜索引擎用户行为研究	124
7.4.2	其他在线搜索系统用户行为研究	125
7.4.3	以特定搜索行为为中心的研究	127
7.4.4	以特定搜索用户群体为中心的研究	128
7.4.5	以查询式为中心的研究	128

7.4.6	以搜索会话为中心的研究	129
7.4.7	查询意图研究	130
7.4.8	搜索系统功能研究	130
7.4.9	搜索日志分析研究现状总结	131
7.5	本章小结	133
第 8 章	点击流数据分析方法	134
8.1	点击流数据分析研究现状	134
8.1.1	以电子商务为背景的研究	135
8.1.2	以社交媒体为背景的研究	138
8.1.3	以在线学习为背景的研究	141
8.1.4	以信息门户为背景的研究	142
8.1.5	以健康医疗为背景的研究	142
8.1.6	以搜索引擎为背景的研究	142
8.1.7	以学术交流为背景的研究	143
8.1.8	点击流数据的可视化分析	143
8.1.9	点击流数据分析研究现状总结	156
8.2	点击流数据分析过程	161
8.3	点击流数据分析方法探索	162
8.4	点击流数据分析方法通用框架构建	165
8.4.1	点击流数据分析层次总览	165
8.4.2	点击流数据的足迹层分析	166
8.4.3	点击流数据的移动层分析	167
8.4.4	点击流数据的路径层分析	169
8.5	本章小结	171
第 9 章	学术图书馆 OPAC 系统用户信息搜寻行为研究	172
9.1	学术图书馆 OPAC 系统概述	172
9.2	研究问题与研究意义	173
9.3	理论基础与相关研究	174
9.3.1	探寻式搜索理论基础	174
9.3.2	学术图书馆 OPAC 系统用户行为相关研究	177
9.4	研究背景——武汉大学图书馆 OPAC	178
9.5	武汉大学图书馆 OPAC 用户信息搜寻行为的点击流数据分析	180
9.5.1	点击流数据准备	180
9.5.2	点击流数据分析	182

9.6 学术图书馆 OPAC 用户信息行为的搜索日志分析	188
9.6.1 搜索日志数据准备	188
9.6.2 搜索日志数据分析	188
9.7 研究讨论	196
9.7.1 独立搜索界面利用率低	196
9.7.2 探寻式搜索极少发生	197
9.7.3 搜索行为遵循“最小努力法则”	197
9.7.4 个人图书馆的发展	198
9.8 本章小结	199
第 10 章 社会性图书馆系统用户信息搜寻行为研究	200
10.1 社会性图书馆系统概述	200
10.2 研究问题与研究意义	201
10.3 理论基础与相关研究	202
10.3.1 信息搜寻方式理论基础	202
10.3.2 社会性标签系统相关研究	204
10.4 研究背景——豆瓣网	205
10.5 豆瓣网用户信息搜寻行为的点击流数据分析	206
10.5.1 点击流数据准备	206
10.5.2 点击流数据分析	208
10.6 豆瓣网用户信息搜寻行为的问卷调查研究	220
10.6.1 问卷设计与调查过程	220
10.6.2 问卷调查研究结果	221
10.7 研究结果讨论	227
10.7.1 主页资源推荐的重要性	227
10.7.2 社会编目者与信息搜寻者的双重身份	228
10.7.3 信息探索的主导地位	229
10.7.4 信息搜寻路径的多样化特征	230
10.8 本章小结	230
参考文献	232
附录 A 武汉大学图书馆 OPAC 系统页面编码体系	263
附录 B 豆瓣网页面编码体系	265
附录 C 豆瓣网用户调查问卷	267

第 1 章 导 论

1.1 在线信息搜寻活动

信息搜寻是人类日常生活和工作中的一项基本活动。为了完成任务、制定决策、解决问题、甚至是满足好奇心、休闲娱乐,人们通过与人工的或数字化的信息系统发生交互来查找信息,从而改变既有的知识状态^[1,2]。早在 19 世纪,人们就习惯从图书馆获取信息,包括图书、期刊、地图、绘画等。馆藏资源的存储和检索离不开规范的编目工作,图书馆采用正式的分类体系和主题词表对物理资源进行组织,电子资源普及后则由元数据方案进行描述^[3]。互联网的出现使得人类信息搜寻活动从线下转移到线上,随着网络环境从 Web 1.0 向 Web 2.0 发展,人们的在线信息搜寻行为也在不断发生重要改变。

在 Web 1.0 时代的早期,用户主要使用网络浏览器访问特定网站,网站中的页面基本上都是静态的,提供文字、链接、图片等内容,这些内容大多由专门人员负责编辑并发布。用户利用页面之间的超链接实现了网络导航,从而可以对信息进行定位;但是由于用户只能在页面上读取信息,信息的流动是单向的。后来,由后台数据库支撑的动态内容逐渐增多,用户与网站实现了双向交流,他们不仅能够读取信息,还能够提交信息,网站在某种意义上变成了网络应用。随着网络信息资源在内容和形式上的多样性不断增加,用户在这样一个异质空间中的信息搜寻活动也呈现多样化的趋势。以 Yahoo! Directory 为代表的人工编制网络层级目录成了用户查找网站和信息的入口,为其发现感兴趣的资源提供了浏览的方式。随后,以 Google 为代表的网络搜索引擎迅速崛起,基于关键词匹配的搜索成为在线信息搜寻的主流方式,并一直延续至今。此外,电子邮件、聊天室和即时通讯应用等也为人们提供了直接信息交流的工具。

进入 Web 2.0 时代以来,网络表现出更为显著的复杂性和动态性,因为用户不仅可以创建、存储、分享自己的信息,而且可以与其他用户发生多向交流。这一切都得益于各种网络技术的进步,在很大程度上提升了在线信息搜寻体验。社交媒体平台允许用户充分参与进来,成为网络内容的贡献者:维基用户可以就共同感兴趣的主题开展协作编辑,博客用户则可以自由发布文章,在表达观点的同时也获得读者的反馈;社会网络服务为用户与联系人分享文字、图片、音视频等信息提供了便利,Facebook、Twitter 都属于通用性的社会性网络服务,而有的服务则专门针对特

定的主题,例如,图片分享(Flickr)和职场社交(LinkedIn)。定制服务技术赋予用户控制网站内容的权限,他们可以依据个人需要订阅新闻和博客,添加工具或服务并按照自己喜欢的方式来安排;另一方面,个性化服务技术能够根据用户的个人资料或活动推荐相关信息。标签和话题标签是这一时代所特有的信息组织方式,在缺乏编目权威的分散式信息空间,用户贡献的非结构化标签为信息资源的访问与获取提供了新的渠道。

在线信息环境在技术的推动下不断演进,为人类信息搜寻活动带来了更多的可能性。这种趋势也将一直持续下去,因为网络正朝着整合空间的方向发展,移动设备、可穿戴设备、传感器和物联网、甚至是人工智能将成为下一代网络的基本特征。在这一背景下,在线信息搜寻也将面临更多的挑战,由信息过载造成的信息焦虑将变得更加严重。信息过载是信息输入超过人类信息处理能力时的状态,如果人们无法避免这种状态就会产生压力或焦虑,这是面对大量信息却无计可施的无力感。因此,在互联网产业努力实现技术飞跃的同时,也应该充分考虑到用户的真实体验,这正是在线信息搜寻行为研究的目的之所在。

一直以来,图书情报学都相对稳定地保持着以信息搜寻与检索(Information Seeking and Retrieval)和信息计量(Informetrics)为两大核心研究领域的构成。其中,信息搜寻与检索领域又可以进一步划分为以系统为导向的信息检索算法研究和以用户为导向的信息搜寻行为研究。进入20世纪90年代以后,这两个研究方向不断交融,有关用户行为的研究结论为系统开发奠定了必要的基础。与此同时,用户信息搜寻行为研究本身也逐步从理论上升到了实证,并且从用户与检索系统的交互扩展到更为广泛的信息传播与获取^[4]。以下本书将首先简要回顾信息搜寻行为研究的发展历程。

1.2 信息搜寻行为研究的发展历程

1.2.1 萌芽阶段(20世纪40、50年代)

目前对于信息搜寻行为领域的起源尚未形成定论。Wilson认为最早的信息行为研究可以追溯到1916年的一项公共图书馆读者阅读习惯研究^[5],而Poole则认为应该是1902年的一项图书馆馆藏使用情况研究^[6]。尽管20世纪初就出现了零星的相关研究,学者们普遍认同信息行为领域到40年代才逐渐进入萌芽阶段,当时两个重要的社会因素引起了人们对信息行为的关注。

一方面,美国政府开始重视公共图书馆的建设,由于需要通过地方税收提供资金支持,针对图书馆开展调查是有必要的,调查问题大多涉及社区对图书馆服务的需求。另一方面,第二次世界大战开启了人类的“大科学”时代,各国政府看到了

信息的散播与传递对于科学家和工程师获得新研究发现的重要意义,而“大科学”带来的海量报告文献在战后被逐步公开,人们可以获取的信息量猛增。

1948年,以改进科学文献收集、索引和发布现有方法为主旨的“皇家协会科学信息会议”(Royal Society Scientific Information Conference)在英国伦敦召开,会上 Bernal 教授的 *Preliminary analysis of pilot questionnaire on the use of scientific literature* 一文报告了他利用多种形式的问卷了解科学家的文献阅读行为,可能是领域内最早的“用户研究”^[7]。

1.2.2 起步阶段(20世纪60、70年代)

1959年,“科学信息国际会议”(International Conference on Scientific Information)在美国华盛顿召开,延续了1948年会议的主题。此次会议上出现了更多与信息使用相关的论文,主要考察科学家们对期刊、图书馆、专利、摘要和索引服务等文献资源的使用情况,旨在为信息服务和检索系统的开发提供参考。这些相关论文基本上都属于用户研究,它们所采取的研究方法以自填问卷和访谈表为主,因而其研究数据和结果都是以定量形式呈现的。

在信息行为领域的起步阶段,欧美国家的图书馆服务进一步发展,人们依然热衷于图书馆调查。1964年英国颁布“公共图书馆和博物馆法案”(Public Libraries and Museums Act),从而引发了一系列的公共图书馆使用研究,其中 Clement 的研究非常具有代表性,涉及33个公共参考图书馆,调查了近30,000名读者的实地、电话、电报、邮寄查询情况,发现他们中70%的人都找到了所需信息^[8]。除了公共图书馆外,学术图书馆也逐步引起了人们的注意,后者在大学师生的研究和学习中扮演了重要角色,所以研究人员试图了解他们对大学图书馆服务的满意度, Ford 曾对该主题相关研究进行了综述^[9]。总的来说,这一时期的信息行为研究多以图书馆为背景,核心的研究问题是人们为什么以及如何使用图书馆。由于大部分研究是基于定量的问卷和访谈方法,它们表现出明显的客观性,即客观地描述图书馆的使用情况,而忽视了信息行为研究理论基础的构建。

1965年,美国文献机构(American Documentation Institute)和国家科学基金(National Science Foundation)联合创立了综述刊物《信息科学与技术年度评论》(Annual Review of Information Science and Technology)。Menzel 和 Paisley 分别于1966年和1968年在该期刊上发表了题为 *Information needs and uses in science and technology* 和 *Information needs and uses* 的综述,反映了当时科学家利用科学信息的行为、习惯、经历和需求以及他们的信息收集和传播行为^[10,11],至今仍被引用。

到了20世纪70年代,用户研究的对象逐渐从科学技术领域转变到社会科学领域。英国的两项研究,即社会科学的信息要求(Information Requirements of the Social Sciences,缩写为 INFROSS)和社会服务部门的信息需求和使用(Information Needs

and Uses in Social Services Departments, 缩写为 INISS), 就充分显示出了这种趋势^[12]。前者主要利用问卷调查了 2,000 多名社会科学家; 而后者则对社会服务部门的社会工作者和管理者开展了长期观察, 该过程中不仅采集到了大量的定量数据, 也采集到了定性数据, 这也使得人们开始意识到将定性方法应用于信息搜寻行为研究的必要性。

1.2.3 形成性发展阶段 (20 世纪 80、90 年代)

20 世纪 80、90 年代是信息行为领域的形成性发展阶段, 绝大多数高被引的信息相关模型和理论都出自那个时期, 同时新的研究方法论观点不断引入, 为其在图书馆与信息科学中抢占核心地位奠定了基础。从本质上讲, 前一阶段的用户研究是以信息系统或服务为中心的, 比如调查用户使用图书馆的情况是为了了解图书馆作为信息传播渠道的有效性, 研究结论可以为其发展实践提供依据, 因此以往的研究具有较强的实用性。而形成性发展阶段的研究核心转移到用户自身上来了, 研究人员关心的是人类作为信息搜寻者的特征, 突出了研究的学术性。

这 20 年内出现了许多反映人类信息行为各个方面的经典理论模型, 包括 Wilson 的信息行为全局模型^[13]、Belkin 等的“不规则知识状态”(Anomalous State of Knowledge, ASK) 概念^[14]、Kuhlthau 的信息搜索过程模型^[15]、Bates 的演进式搜索采莓模型^[16]等, 后来人们在不同的研究背景下检验了这些模型。该时期的实证研究则形成了针对细分人群开展研究的特点, 研究对象通常是从事特定职业的某一类用户, 如学者、医生、律师、艺术家等, 或是扮演特定社会角色的人, 如消费者、病人、学生等。定性方法在信息行为研究中得到进一步推广, 同时考虑到信息行为持续发生和演进的动态特征, 在时间上具有延展性的纵向定性方法受到了青睐, Fidel 的搜索关键词选择研究^[17]、Case 的人文学者信息搜寻动机和方法研究^[18]、Qiu 的超文本导航路径研究^[19]等都采用此类方法。随着越来越多的研究人员开始进入该领域, 相关论文产出数量猛增, 而且逐步占据了图书馆与信息科学的核心期刊, 包括《美国信息科学与技术协会杂志》(*Journal of the American Society for Information Science and Technology*)、《信息研究》(*Information Research*)、《信息处理与管理》(*Information Processing and Management*)、《文献杂志》(*Journal of Documentation*)、《学术图书馆学杂志》(*Journal of Academic Librarianship*)、《图书馆与信息科学研究》(*Library and Information Science Research*) 等。

90 年代中后期, 信息行为领域内发生了三个具有重要意义的里程碑事件。

(1) 1995 年, 聚焦于信息行为研究的国际学术期刊《信息研究》创刊, 一直以来由英国谢菲尔德大学 Wilson 教授担任主编, 他本人是信息行为领域的著名学者, 曾贡献了很多重要的研究成果, 作者 h 指数高达 45。《信息研究》是一本同行评审的开放存取期刊, 围绕信息行为、信息搜寻行为、信息需求、信息搜索、信息检索、

信息使用等核心主题刊登论文,并广泛涉及信息科学、信息管理、信息技能、健康信息、互联网、教育等主题。经过20多年的发展,该期刊的影响力已经遍及世界各国,并且列入社会科学引文索引(Social Science Citation Index)、期刊引文报告(Journal Citation Reports)、图书馆与信息科学文摘(Library and Information Science Abstracts)以及谷歌学术(Google Scholar)。

(2) 1996年,信息行为专题国际会议(Information Seeking in Context,简称ISIC)首次在芬兰坦佩雷召开,Wilson教授也是会议发起人之一,目前由芬兰坦佩雷大学的Reijo Savolainen教授担任主席。该会议每隔一年召开一次,至今已经成功举办11次,地点主要选在欧洲城市。跨学科交融性是ISIC最主要的特点,会议将来自信息科学、信息管理、社会性、心理学、信息系统、计算机科学的学者汇聚到一起,共同探讨用户需求和需求满足之间的关系,具体的讨论主题包括信息搜寻和搜索理论与模型、研究方式和方法论、特定背景(如医疗、教育、商业、政府等)下的信息搜寻和使用分享、虚拟社交网络中的信息搜寻与搜索、日常生活信息搜寻以及协作式信息搜寻等。

(3) 1999年,美国信息科学与技术协会(American Society of Information Science and Technology, ASIS&T)成立了“信息需求、搜寻与使用”特别兴趣小组(Special Interest Group on Information Needs, Seeking & USE, SIG USE)。特别兴趣小组为研究兴趣相似的ASIS&T会员提供了了解领域发展、交流想法的机会,SIG USE主要关注用户与信息交互过程中的行为、认知活动和情感状态,包括信息需求的产生与识别以及信息的搜寻、分享、管理和交流等,旨在通过各种小组内部活动推动信息相关人类行为的研究,尤其鼓励将研究成果应用于信息系统的设计与开发。

1.2.4 爆炸性发展阶段(2000年以后)

这一阶段依然推崇以用户为中心的研究,以上提到的按职业和社会角色对用户进行细分的研究范式得到了延续。Julien等对发表于1999~2008年的749篇相关文献进行内容分析发现,常见的研究对象包括学生、学者、专业人员、普通民众等,而有近一半研究的对象为“其他”群体,即网络用户^[20]。20世纪90年代是互联网飞速发展并大规模普及的重要十年,其间一批网络搜索引擎陆续涌现,如Excite、Lycos、Altavista、Google等,成为人们在互联网上搜寻信息的主要工具,这也使得信息搜寻行为研究逐步聚焦于人类搜索行为。2000年左右,围绕用户与搜索系统的交互开展的搜索日志分析(Search Log Analysis)出现了,随后迅速增长。这类研究利用来自搜索系统的日志文件获得用户的真实行为数据,通过搜索会话、查询式和关键词层次的分析来了解用户搜索行为的外在特征,并进而回答各种研究问题。搜索日志分析的引入标志着信息搜寻行为领域在研究方法上的突破,不再局限于传统的介入性方法,而转向对非介入性方法的应用。

除了通用网络搜索引擎外,搜索日志分析还广泛涉及网站、图书馆 OPAC、数字图书馆等的搜索系统以及元搜索引擎、多媒体搜索引擎、移动搜索系统、垂直搜索系统等,为网络搜索技术的进步和搜索交互体验的提升提供了有效的方向指引。然而这还不足以覆盖人类搜索行为的全貌,因为搜索日志中的客观行为数据无法揭示用户的主观动因和搜索情境的影响。因此,与搜索日志分析并行发展的还有用户行为影响因素研究,通过问卷、访谈、实验、观察等方法了解人们的人口统计信息、搜索技能、领域知识、认知方式和情感特征等,并考察他们所处的信息环境和所面临的搜索任务,基于系统的研究设计探讨这些方面对用户搜索行为过程和最终结果的塑造作用,从而为用户信息素养的培养和搜索系统的功能与界面设计带来许多有用的启示。

随着 Web2.0 时代的来临,博客、微博、维基、社会网络服务、社会性标签系统、虚拟世界等社交媒体兴起,成为互联网发展进程中继搜索引擎之后的又一个里程碑。社交媒体首先是一个参与和协作的平台,依赖用户贡献内容,并利用“集体智慧”(Collective Intelligence)使内容的价值最大化,来自于社会学、心理学、传播学、计算机科学等学科的研究人员针对用户的信息分享、交流、学习、决策等行为进行了探讨^[21]。同时,社交媒体也为人们获取信息创造了一个开放、灵活、多样的环境,以此为背景的信息搜寻研究将焦点转移到用户的导航行为上,试图了解并提高他们通过资源推荐、社交网络、标签云等社交媒体所独有的功能实现信息探索的效果^[22]。相反,由于词汇问题和基本层次问题的存在,社交媒体中的已知条目搜索效率不再令人满意^[23,24]。从近几年的 ISIC 会议主题也可以看出,社交媒体环境下的信息行为研究已经成为信息搜寻领域关注的重要内容,相关论文广泛涉及社交媒体用户的信息需求、信息搜寻行为、信息吸收与利用行为^[25-27]等。

2010 年之后,信息行为领域的研究变得更为多样化,研究人员逐渐意识到人类信息偶遇、信息规避行为的普遍性。信息偶遇概念的提出可以追溯到 20 世纪 90 年代中期^[28],但是由于偶遇现象不如主动性高、方向性强的搜索行为那样易于描述、便于捕捉,以至于相关研究在很长一段时间内推进缓慢,直至近几年才取得了一些重要进展。研究人员不仅对信息偶遇的概念内涵进行了深入的探讨,并且在研究方法上做出了各种各样的新尝试,从而不断改进、完善了信息偶遇过程模型^[29-32],同时从用户、信息和环境等多个维度探讨了影响偶遇发生概率的诸多因素^[33-38]。在信息爆炸的网络时代,信息规避现象也随处可见,并且多发生于信息偶遇的背景下,甚至已经成为人们应对信息过载的一种策略。然而信息行为领域对信息规避的研究仍处于起步阶段^[39],尽管可以从有些经典的信息搜寻行为模型中找到一定的线索^[13,40],但是相关实证研究还相当缺乏,规避行为的分类和成因尚不清晰。作为人类信息行为的重要组成部分,信息偶遇和信息规避将成为信息行为领域极具发展前景的研究方向。