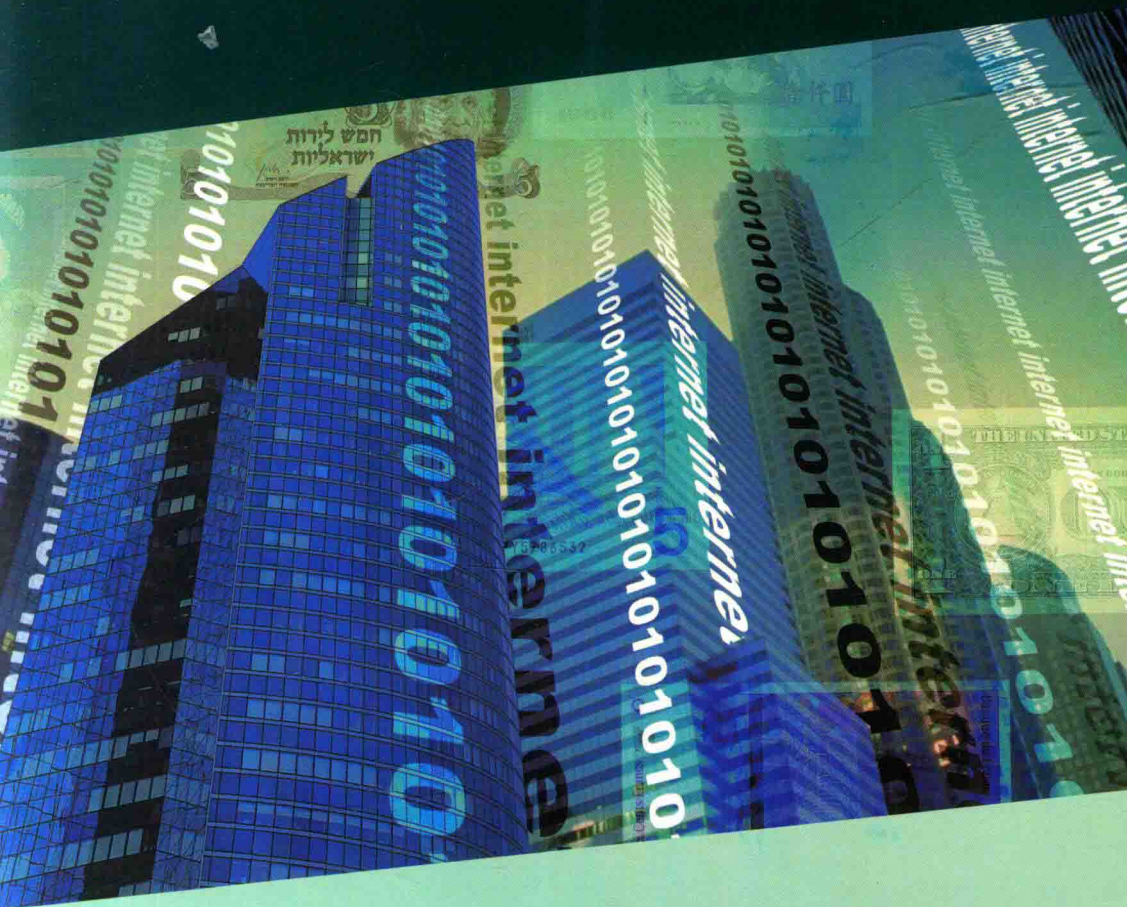




微博信息内容 质量评价及影响分析

莫祖英 著

世界图书出版公司

教育部人文社会科学重点研究基地重大项目

“数字信息资源的质量控制与管理研究”(12JJD820008)

微博信息内容质量评价及影响分析

莫祖英 著

世界图书出版公司

上海·西安·北京·广州

图书在版编目(CIP)数据

微博信息内容质量评价及影响分析/莫祖英著.

—上海:上海世界图书出版公司,2015.9

ISBN 978-7-5100-7244-4

I. ①微… II. ①莫… III. ①互联网络-传播媒介-研究 IV. ①G206.2

中国版本图书馆 CIP 数据核字(2015)第 206535 号

微博信息内容质量评价及影响分析

著 者 莫祖英

出 版 人 陆 琦

责任编辑 姜海涛

装帧设计 刘晓灿

责任校对 石佳达

出版发行 上海世界图书出版公司

地 址 上海市广中路 88 号

电 话 021-36357930

邮政编码 200083

经 销 各地新华书店

印 刷 上海市印刷七厂有限公司

开 本 787×1092 1/16

印 张 12.75

字 数 165 000

版 次 2015 年 9 月第 1 版

印 次 2015 年 9 月第 1 次印刷

书 号 978-7-5100-7244-4/G·445

定 价 35.00 元

www.wpcsh.com.cn

www.wpcsh.com

如发现印装质量问题

请与印刷厂联系 021-59110729

目 录

1 绪 论	1
1.1 选题背景与研究意义	2
1.1.1 选题背景	2
1.1.2 研究意义	4
1.2 国内外研究现状综述	5
1.2.1 微博信息研究	5
1.2.2 信息质量研究	12
1.2.3 微博信息质量研究	22
1.3 研究内容、方法和创新	27
1.3.1 研究内容	27
1.3.2 研究方法	28
1.3.3 创新之处	29
2 相关概念及理论	31
2.1 信息质量	31
2.1.1 信息质量的定义	31
2.1.2 信息质量维度的划分	34
2.2 信息评价理论	39

2.2.1	定性评价	39
2.2.2	定量评价	40
2.2.3	综合评价	41
2.3	内容分析法	42
2.3.1	内容分析法的定义	42
2.3.2	内容分析法的实施过程	44
2.3.3	内容分析法的应用模型	45
3	微博信息内容质量评价模型构建及影响分析	47
3.1	微博信息内容质量的界定	47
3.2	微博信息质量描述指标分析	49
3.2.1	元信息质量指标	49
3.2.2	页面信息质量指标	52
3.3	评价模型的构建	53
3.3.1	权重设置	53
3.3.2	模型构建	56
3.4	信息内容质量对微博利用的影响分析	57
4	微博信息内容质量评价及特征分析	59
4.1	数据采集	59
4.2	微博信息内容质量分析与评价	62
4.2.1	科技用户微博的信息内容质量分析与评价	62
4.2.2	公安用户微博的信息内容质量分析与评价	72
4.2.3	房地产用户微博的信息内容质量分析与评价	79
4.2.4	环保用户微博的信息内容质量分析与评价	85
4.3	微博信息内容质量特征分析	91

4.3.1	微博信息内容质量的量化评估	92
4.3.2	不同领域用户微博页面信息内容质量特征分析	100
5	信息内容质量在不同用户群体中的差异性分析	103
5.1	差异性检验	103
5.2	不同领域用户的微博信息内容质量差异性分析	104
5.2.1	科技用户与其他用户的微博信息质量差异性分析	104
5.2.2	房地产用户与其他用户的微博信息质量差异性分析	111
5.2.3	公安用户与环保用户的微博信息质量差异性分析	115
5.3	对比分析与总结	117
6	信息内容质量对微博转发的影响分析	119
6.1	相关性分析	119
6.2	信息内容质量对微博转发影响的实证分析	121
6.2.1	科技领域用户微博转发率的影响分析	121
6.2.2	房地产领域用户微博转发率的影响分析	129
6.2.3	公安领域用户微博转发率的影响分析	135
6.2.4	环保领域用户微博转发率的影响分析	140
6.3	对比分析与总结	145
7	信息内容质量对微博评论的影响分析	149
7.1	微博评论的影响因素分析	149
7.2	信息内容质量对微博评论影响的实证分析	150
7.2.1	科技用户页面信息质量对微博评论的影响分析	150
7.2.2	房地产用户页面信息质量对微博评论的影响分析	157
7.2.3	公安用户页面信息质量对微博评论的影响分析	162
7.2.4	环保用户页面信息质量对微博评论的影响分析	167

7.3 对比分析与总结 172

8 结论与展望 174

8.1 研究结论与启示 174

8.2 研究不足与展望 175

中外文参考文献 176

附录 184

附录 A 微博信息内容质量评价指标重要性调查表 184

附录 B 攻博期间发表的科研成果目录 187

致谢 189

图表目录

图 2-1 信息质量维度的层次结构	35
图 3-1 元信息质量评价模型	56
表 2-1 标记信息质量维度的 PSP/IQ 模型	36
表 3-1 评价指标的绝对重要度	54
表 3-2 判断矩阵 1	55
表 3-3 判断矩阵 2	55
表 3-4 元信息质量描述指标的权重	56
表 4-1 微博数据采集字段及说明	61
表 4-2 微博信息内容的质量特征	61
表 4-3 科技类用户微博页面信息的文本信息量	63
表 4-4 科技类微博页面信息的超链接信息量	64
表 4-5 科技类微博页面信息内容的类型分布	65
表 4-6 科技类微博页面信息中“@”出现的频率	69
表 4-7 科技类微博用户页面的信息热度	70
表 4-8 科技类微博页面信息中标题的平均数量	71
表 4-9 公安类用户微博页面信息中的文本信息量	72
表 4-10 公安类微博页面信息的超链接总数量	73

表 4-11	公安类微博页面信息内容的类型分布	74
表 4-12	公安类微博页面信息中“@”出现的频率	76
表 4-13	公安类微博用户页面的信息热度	77
表 4-14	公安类用户微博页面信息中标题的平均数量	78
表 4-15	房地产类用户微博页面信息的文本信息量	79
表 4-16	房地产类微博页面信息的超链接信息量	80
表 4-17	房地产类微博页面信息内容的类型分布	80
表 4-18	房地产类微博页面信息中“@”出现的频率	83
表 4-19	房地产类微博用户页面的信息热度	84
表 4-20	房地产类微博页面信息中标题的平均数量	85
表 4-21	环保类用户页面信息中的文本信息量	86
表 4-22	环保类用户页面信息中的超链接信息量	87
表 4-23	环保类微博页面信息内容的类型分布	87
表 4-24	环保类微博页面信息中“@”出现的频率	89
表 4-25	环保类微博用户页面的信息热度	90
表 4-26	环保类微博页面信息中标题的平均数量	91
表 4-27	微博用户页面信息内容类型量化得分(部分数据)	93
表 4-28	归一化处理后的质量指标值及页面信息质量值 (部分数据)	95
表 4-29	微博用户页面信息内容质量值	98
表 4-30	四大领域用户微博页面信息内容质量描述指标平均值	101
表 5-1	科技用户与房地产用户微博页面信息内容质量指标统计量	105
表 5-2	科技用户与房地产用户微博页面信息内容质量指标 差异性检验	105

表 5-3	科技用户与公安用户微博页面信息内容质量指标统计量	107
表 5-4	科技用户与公安用户微博页面信息内容质量指标 差异性检验	108
表 5-5	科技用户与环保用户微博页面信息内容质量指标统计量	109
表 5-6	科技用户与环保用户微博页面信息内容质量指标 差异性检验	110
表 5-7	房地产用户与公安用户微博页面信息内容质量指标统计量	111
表 5-8	房地产用户与公安用户微博页面信息内容质量指标 差异性检验	112
表 5-9	房地产用户与环保用户微博页面信息内容质量指标统计量	113
表 5-10	房地产用户与环保用户微博页面信息内容质量指标 差异性检验	114
表 5-11	公安用户与环保用户微博页面信息内容质量指标统计量	115
表 5-12	公安用户与环保用户微博页面信息内容质量指标 差异性检验	116
表 6-1	相关系数的判别规则	120
表 6-2	某一用户中信息量与微博转发率之间的相关性	122
表 6-3	科技类用户微博信息量与转发率之间的相关性	122
表 6-4	科技类用户微博信息类型与转发率之间的相关性	124
表 6-5	科技用户微博信息真实性与转发率之间的相关关系	125
表 6-6	科技类用户微博信息热度与转发率之间的相关关系	126
表 6-7	科技类用户微博信息编辑质量与转发率之间相关关系	127
表 6-8	房地产用户微博信息量与转发率之间的相关关系	130
表 6-9	房地产用户微博信息内容类型与转发率之间的相关关系	131

表 6-10 房地产用户微博信息真实性与转发率之间的相关关系 132

表 6-11 房地产用户微博信息热度与转发率之间的相关关系 133

表 6-12 房地产用户微博信息编辑质量与转发率之间的相关关系 133

表 6-13 公安用户微博页面中信息量与微博转发之间的相关关系 135

表 6-14 公安用户微博页面中信息内容类型与转发率之间的
相关关系 137

表 6-15 公安用户微博页面中信息真实性与转发率之间的
相关关系 138

表 6-16 公安用户微博页面中信息热度与转发率之间的相关关系 138

表 6-17 公安用户微博页面中信息编辑质量与转发率之间的
相关关系 139

表 6-18 环保用户微博页面信息量与转发率之间的相关关系 141

表 6-19 环保用户微博页面信息类型与转发率之间的相关关系 142

表 6-20 环保用户微博页面信息真实性与转发率之间的相关关系 143

表 6-21 环保用户微博页面信息热度与转发率之间的相关关系 144

表 6-22 环保用户微博页面信息编辑质量与转发率之间的
相关关系 144

表 6-23 四大领域用户群体微博信息质量指标对转发率的
影响作用 146

表 7-1 科技用户微博页面信息量与评论数之间的相关关系 151

表 7-2 科技类用户微博页面信息类型与评论数之间的相关关系 ... 153

表 7-3 科技类用户微博页面信息真实性与评论数之间的相关关系 ... 153

表 7-4 科技类用户微博页面信息热度与评论数之间的相关关系 ... 154

表 7-5 科技类用户微博页面信息编辑质量与评论数之间的

相关关系	155
表 7-6 房地产用户微博信息量与评论数之间的相关关系	157
表 7-7 房地产用户微博信息内容类型与评论数之间的相关关系 ...	159
表 7-8 房地产用户微博信息真实性与评论数之间的相关关系	159
表 7-9 房地产用户微博信息热度与评论数之间的相关关系	160
表 7-10 房地产用户微博信息编辑质量与评论数之间的相关关系	161
表 7-11 公安用户微博信息量与评论数之间的相关关系	162
表 7-12 公安用户微博信息内容类型与评论数之间的相关关系	164
表 7-13 公安用户微博信息真实性与评论数之间的相关关系	164
表 7-14 公安用户微博信息热度与评论数之间的相关关系	165
表 7-15 公安用户微博信息编辑质量与评论数之间的相关关系	166
表 7-16 环保用户微博信息量与评论数之间的相关关系	168
表 7-17 环保用户微博信息内容类型与评论数之间的相关关系	169
表 7-18 环保用户微博信息真实性与评论数之间的相关关系	169
表 7-19 环保用户微博信息热度与评论数之间的相关关系	170
表 7-20 环保用户微博信息编辑质量与评论数之间的相关关系	171
表 7-21 四大领域用户群体微博信息质量指标对微博评论的 影响作用	172

1 绪论

20 世纪 90 年代以来,信息技术不断创新,信息产业持续发展,信息网络广泛普及,信息化成为全球经济社会发展的显著特征,信息资源日益成为重要生产要素、无形资产和社会财富。信息资源是信息经济时代重要的国家战略资源,是实现经济与社会可持续发展的基础条件。随着现代信息技术迅速发展,特别是网络环境的形成,信息的生产、存储和传递的方式发生了革命性的变化,互联网信息资源以传统信息资源难以比拟的优势逐渐成为信息资源的主体。根据中国互联网络信息中心(CNNIC)2013 年 7 月发布的统计报告显示,截至 2013 年 6 月底,我国网站数量为 294 万,增长率为 9.6%,网页数量为 1227 亿个,增长率为 41.7%;2012 年单个网站的平均网页数和单个网页的平均字节数均维持增长,反映出我国互联网上的信息内容更为丰富。与此同时,信息资源结构也发生了较大变化,非结构化信息和半结构化信息大量出现,与早期基于信息系统和数据库开发的结构化数据为主的模式不同,非结构化信息资源泛滥,其质量参差不齐,大量低劣信息、假信息、伪信息无控制传播,给网络信息资源管理带来极大挑战,严重影响了网络信息资源的有效开发与利用。

随着 Web2.0 技术的发展,微博成为互联网的一大主流应用模式。它最早产生于 2006 年,2007 年开始引入中国,经历了 2010—2011 年爆发式增长后,从 2012 年开始进入了一个相对平稳的增长期。微博集社交、休闲、娱乐、营销于一身,满足了人们在碎片化时间里弱关系、低强度的上网沟通需求,吸引了大批重度用户。目前,微博已成为网民获取信息的主要途径之一,越来越多的机构、公众人物通过微博来发布与传播信息。微博在蓬勃发展的同时,也产生了很多问题。由于“把关人”的缺失、信息的碎片化特征、

用户水平良莠不齐,微博中充斥着种种乱象。庞大的用户群与实时传播功能造成微博信息泛滥、谣言四起,信息发布的便捷与转发功能造成大量的信息冗余,信息发布中“把关人”角色的缺失使失真信息、琐碎信息和庸俗信息充斥着整个微博。同时,微博营销的快速发展引发了一系列不规范操作,甚至违法犯罪行为。根据 CNNIC 对用户在使用微博中遇到的问题进行的调查显示,目前微博承载的信息内容不能满足用户的质量需求,主要表现在信息量过大、有效价值不高且获取难等方面。微博信息内容质量问题已成为影响微博平台健康发展、网络信息资源有效利用的重要问题,微博信息资源质量治理迫在眉睫。因此,开展微博信息内容质量评价,分析信息质量对用户利用的影响,为控制与改进微博信息质量、提高用户对微博信息的利用率提供有益的策略与思路,也就成了本书研究的核心内容。

1.1 选题背景与研究意义

微博是 Web2.0 应用中的一大主流模式,其开放性、交互性、便捷性等特点受到众多用户的青睐,成为移动互联网时代最具发展潜力的应用之一。微博平台快速发展的同时,各种乱象丛生,信息质量问题严重,为其发展带来了困扰。微博信息内容质量评价及影响分析有利于提高微博信息质量,促进微博平台的健康发展。

1.1.1 选题背景

微博平台快速发展的同时,也产生了许多问题,尤其是微博信息内容方面的质量问题,如虚假信息、低俗信息充斥着整个微博空间,为用户信息利用带来了困难与障碍。

(1) 微博平台蓬勃发展,成为网民获取信息的主要来源

随着 Web2.0 技术的发展,微博成为互联网的一大主流应用模式。所谓 Web2.0,它是相对于 Web1.0 的新的互联网应用模式,其主要特点是用户

生成内容、注重用户的交互作用。微博作为开放式社交服务的一种,允许用户通过电脑、手机等多种平台撰写、评价与分享信息,内容限制在 140 个字以内,具有开放性、便捷性、及时性、互动性、群聚性等特点。微博信息内容的“碎片化”特征使其用户更加平民化,多平台发布的便捷性特征使其信息内容更新及时、快速,点对面的传播方式使其信息传播速度快、范围广,基于用户关系的信息分享与传播模式促进了用户间的交流频率和交流速度。微博的这些特征显现出强大的信息传播能力和良好的社交功能,很快赢得了社会的广泛认可。

微博最早产生于 2006 年,其形式是美国的 Twitter,这是最早也是最著名的微博产品。2007 年微博开始引入中国。2009 年,随着新浪微博平台的测试上线,进入全面发展时期。2010 年被称为“中国微博元年”,从这一年起,微博在我国呈井喷式发展。根据中国互联网信息中心(CNNIC)调查数据显示,截至 2013 年 6 月底,我国微博用户规模为 3.31 亿,占整体网民规模的 56%,其中手机微博用户规模达到 2.30 亿,占手机网民的 49.5%。庞大的用户规模凸显了其在网络舆论传播的中心地位,微博正在重塑社会舆论生产与传播机制,而微博用户行为的移动化使其成为移动互联网时代最具发展潜力的产品之一。

微博形式精简,功能强大,支持文字、图片、视频等多媒体信息,信息内容具有较强的关联性和连续性;微博用户间关系的双向性可同时满足多层次社交需求,构建起强关系与弱关系并存的网络;微博的转发分享功能使每个用户都是信息传播者,形成一张庞大且高效的信息传播网络。总之,微博正发展成为互联网中重要的信息来源,其主要作用表现为社会新闻、社交网络、监督平台和社会干预,日益成为网民生活中不可缺少的信息来源媒介。

(2) 微博平台中乱象丛生,信息内容质量问题严重

微博在蓬勃发展的同时,也产生了很多问题。微博营销的日益兴起,增加了大量广告推广内容,且虚假广告大行其道,强制推广和私信广告泛滥,使用户体验大大降低。同时,微博中虚假信息横行,用户难以识别,无形中成为传播者。在每天快速更新的大量微博信息中,用户如何更好地获取高

质量、高价值的微博信息是亟待解决的问题。

1.1.2 研究意义

(1) 理论意义

本研究的理论意义主要表现在三个方面:

一是丰富和发展信息质量理论体系。本研究的目标是实现微博信息内容质量评价,并分析信息质量对用户利用的影响,以挖掘关键影响因素,为实现微博信息质量控制与有效利用提供新思路。达成这一目标,实际上是对信息质量理论在社会化媒体这一新环境下的丰富与发展。

二是弥补微博信息质量研究中的明显疏漏。目前微博信息质量研究主要集中在信息质量评价和信息质量问题分析方面,缺乏对如何提升与改进微博信息质量的研究,本书采用实证方法分析微博信息内容质量对用户利用的影响,为微博信息质量控制与有效利用提供基础与条件,这必然拓宽其研究思路,深化其研究内容,推动其进一步发展。

三是对信息资源管理理论的完善。微博信息是 Web2.0 环境下的重要网络信息资源,也是社会化媒体环境下网络信息资源的重要代表。信息质量评价及影响分析是实现信息资源管理的基础与前提,本书对社会化媒体环境下微博信息质量的影响研究,是对新环境下网络信息资源管理理论的发展与完善。

(2) 实践意义

本研究的实践意义主要表现在两个方面:

一是优化微博信息质量,更好地满足用户信息需求。本书通过对微博信息内容质量对用户利用的影响分析,便于发现关键影响因素,进而实现对微博信息质量的控制与管理,更好地满足微博用户的质量需求。

二是为微博平台的健康发展提供指导与建议。本书研究有利于实现微博信息质量的控制与管理,可为微博平台的健康发展提供指导与建议。

1.2 国内外研究现状综述

微博(microblog)作为近年来兴起的一种基于用户关系的信息分享与交流平台,集成了手机短信、社交网络和 BBS 论坛的优势,具有内容简洁、使用便利、快捷、易于传播和交互性强的特点,日益成为国内外学者们的主要研究对象之一。本书的研究现状综述力求全面且突出重点,首先针对微博信息的研究现状进行综述,然后全面梳理信息质量理论研究现状,最后对微博信息的质量研究现状进行文献梳理,为本书研究提供理论基础,同时为进一步深入研究提供研究线索与思路。

1.2.1 微博信息研究

目前国外学者对微博的研究主要以 Twitter 为例,内容集中在对 Twitter 用户和 Tweets 信息的研究上,研究热点包括微博信息检索研究、微博信息监测与舆情分析、微博行为预测等。芮(Rui)和安德鲁·温斯顿(Andrew Whinston)^①通过监测 521 个活跃的 Twitter 用户,分析了用户对 Twitter 的贡献率。张(Cheong)及文森特·C.S.李(Vincent C. S. Lee)^②将微博用于恐怖信息的研究,根据 2009 年孟买恐怖袭击事件,观察市民在 Twitter 中的反应,并通过智能数据挖掘与可视化,揭露市民对恐怖威胁的潜在反应。安德烈亚斯·C.索南比勒(Andreas C. Sonnenbichler)和克里斯

① Rui H.X., A. Whinston, "Information or Attention? An Empirical Study of User Contribution on Twitter." *Information System E-Business Management*, 2012, No. 10, pp. 309 - 324.

② M. Cheong, V. C. S. Lee, "A Microblogging-Based Approach to Terrorism Informatics: Exploration and Chronicling Civilian Sentiment and Response to Terrorism Events via Twitter." *Information System Front*, 2011, No. 13, pp. 45 - 59.