

语言舆情与 语言政策探索

魏晖 主编

海外借



SINCE 1897

商務印書館
The Commercial Press

国家语委“十二五”科研规划项目（ZDI1259）资助出版



语言舆情研究中心

语言舆情与 语言政策探索

魏晖 主编



SINCE 1897

商务印书馆

The Commercial Press

2016年·北京

图书在版编目(CIP)数据

语言舆情与语言政策探索/魏晖主编. —北京:商务印书馆, 2016

ISBN 978-7-100-12727-1

I. ①语… II. ①魏… III. ①舆论—关系—语言政策—研究 IV. ①C912.63 ②H002

中国版本图书馆 CIP 数据核字(2016)第 275498 号

所有权利保留。

未经许可,不得以任何方式使用。

语言舆情与语言政策探索

魏 晖 主编

商 务 印 书 馆 出 版

(北京王府井大街 36 号 邮政编码 100710)

商 务 印 书 馆 发 行

北京市艺辉印刷有限公司印刷

ISBN 978-7-100-12727-1

2016 年 12 月第 1 版

开本 850×1168 1/32

2016 年 12 月北京第 1 次印刷

印张 8 3/4

定价:27.00 元

前 言

随着我国互联网普及率持续提高、网民队伍进一步扩大和网络传播快速发展,舆情研究尤其网络舆情研究越来越受到学界的关注。不少高等学校、科研机构、媒体纷纷成立舆情研究机构,开展舆情监测分析及相关理论和对策研究,发布年度舆情报告、舆情应对能力评估报告等。国家自然科学基金和社科基金等加大了对网络舆情发生机制、传播模式、管理尤其突发网络舆情事件演化、预警、研判和调控疏导等研究的支持力度。

当前,我国社会语言生活空前活跃,全面准确把握语言生活动态,十分必要而紧迫。然而,处在信息化社会,面对海量信息,往往会茫然困惑,难以判断其真实性,无法适时做出恰当的决策。因此,语言文字舆情的监测、汇集和分析至关重要。

为了全面做好语言文字舆情的监测、汇集、分析与研究,教育部语言文字应用研究所在教育部语言文字信息管理司的支持下,于2009年初成立语言舆情研究中心并创办了《语言文字舆情与动态》和《语言文字舆情月月读》内部电子期刊(2013年合并,更名为《语言舆情扫描》),建立语言舆情库,组织舆情分析会,力图以信息聚合技术为手段,以语言文字社会热点为专题,跟踪分析舆情,把

握舆情状况及趋势,及时提供全面准确的语言文字舆情与动态信息以及相关舆情分析报告,以期“透过语言舆情了解语言生活,透过语言生活思考语言政策”,为政府决策和公众了解社会语言生活做好服务。中心建设伊始,请北京航空航天大学新媒体与艺术设计学院李巧老师为中心设计了标识(LOGO),寓意为:语言文字从“点”开始;语言舆情点滴汇集;语言舆情研究,从一点一滴做起,脚踏实地,力争为国家语言文字事业做出点滴贡献。

语言舆情研究团队 2009 年年初组建后不久,8 月份便恰逢《通用规范汉字表》公开征求意见。8 月 12 日至 9 月 8 日,语言舆情研究团队每天对平面媒体、有声媒体、网络媒体等进行舆情跟踪监测,同时将每天的舆情汇总、提要、分析,编写成《舆情快报》呈送征求意见领导小组办公室,供领导及时了解掌握舆情动态,便于制定或调整相关对策。公开征求意见工作结束后,对舆情进行了汇总、研究,提交了相关资料汇编和舆情分析报告。这是一次大的考验,舆情研究团队克服人手不够、经验不足的困难,圆满完成了任务。2013 年在《通用规范汉字表》发布期间,也每天重复着上述工作。7 年来,尽管遭遇过经费短缺、人手不足等困难,舆情研究中心一直默默坚持着,编辑内刊,研讨舆情热点,撰写舆情分析报告,开展监测方法、语言政策研究,本书收集了自组建以来撰写的部分舆情分析报告和发表的学术论文 20 篇(为读者阅读方便,大致按时间顺序排列,相关内容文章、报告相对集中),希冀对以往的工作做一梳理、总结,通过总结,也期待舆情监测、分析、研究工作更上一层楼。

语言舆情监测研究工作获得了国家社科基金、国家语委科研规划项目等的支持和资助,也得到了教育部相关司局、语言文字应用研究所各位领导和同仁以及应用语言学学界诸多专家的指导、支持和鼓励,在此特致谢忱!最后感谢商务印书馆对本书出版的支持!

目 录

语言舆情监测与社会语言生活	刘靖文 张 挺(1)
2009 年语言文字热点舆情综述	刘靖文(15)
汉字简繁之争的由来及评析	刘靖文(32)
《通用规范汉字表》公开征求意见舆情及分析 报告	刘靖文(43)
2010 年语言生活聚焦	孙曼均(59)
2010 年语言文字舆情盘点	张 挺 魏 晖(73)
汉语不存在危机,国民需要语言教育	孙曼均(83)
网络词语的特点和舆情视角下引发的思考	陈 菲(92)
外语使用中存在的问题与对策	魏 晖(102)
我国国际大都市外语学习和使用状况及政策 分析	魏 晖(111)
互联网环境下语言文字舆情监测与实证 研究	张 挺 魏 晖(124)
当前语言文字舆情特点与走势分析	孙曼均(139)
媒介与语言:网络传播对当代社会语言生活 影响之考察	张 挺 魏 晖(153)

语言舆情热点的社会文化分析及其启示·····	孙曼均(164)
语言舆情视角下和谐语言生活构建策略	
研究·····	张 挺 刘靖文(181)
舆情视角下的汉语能力测试·····	张 平(194)
浅议汉语的发展与规范·····	魏 晖(208)
从字母词使用视角再议汉语的规范·····	魏 晖(217)
方言保护热议·····	谢俊英(228)
舆情视角下语言教育问题分析·····	魏 晖(241)

语言舆情监测与社会语言生活^{*}

刘靖文 张 挺

互联网的迅猛发展带动了中国网民队伍的迅速扩大。2010年1月15日,中国互联网络信息中心(CNNIC)发布了《第25次中国互联网络发展状况统计报告》(以下简称《报告》)。《报告》数据显示,截至2009年12月,我国网民规模已达3.84亿,互联网普及率进一步提升,达到28.9%。较2008年年底增长8600万人,年增长率为28.9%。中国手机网民规模年增加1.2亿,达到2.33亿人,占整体网民的60.8%^[1],我国网民规模、宽带网民数、国家顶级域名注册量(1296万)三项指标仍然稳居世界第一,互联网普及率稳步提升。互联网是完全开放的,每个人都有机会成为网络信息的发布者,每个人都有选择网络信息的自由。据统计,56%的中国网民经常在互联网上发布意见,网民参与和表达的意愿持续高涨。正如胡锦涛总书记所言,互联网已成为思想文化信息的集散地和社会舆论的放大器。互联网也逐渐成为党和政府了解社会民情民意及治国理政的重要新平台之一。因此,在加强互联网信

^{*} 本文原刊于《云南师范大学学报》(哲学社会科学版)2011年第1期,个别处有修改。

息监管的同时,组织力量开展舆情监测与研究,汇集整理和分析信息,对于及时应对网络突发的公共事件、全面掌握社情民意、为党和政府的决策服务,推动改革开放和现代化建设方面的重要作用,日渐发挥和显现出来,更加引起关注和重视。

一、语言舆情监测的必要性

语言(文字)是人类社会最为重要的交际工具和思维工具,是文化的重要组成部分和基本载体。随着信息化、全球化进程的不断加快,语言的地位和价值空前凸现,很多国家都逐渐开始重视语言,一些发达国家将语言战略视为国家战略发展的重要组成部分。我国当前语言生活正快速发展变化,社会语言生活空前活跃,语言生活中各种矛盾凸显,社会需要提供语言服务的类型与方式与日俱增,全面准确把握语言生活动态,十分必要而紧迫。然而,处在信息化社会,面对海量信息,往往会茫然困惑,难以判断其真实性,无法适时做出恰当的决策。因此,语言文字舆情的监测、跟踪及分析至关重要。

具体而言,语言文字舆情监测与分析的研究价值体现在以下几个方面:

1. 透过舆情了解社会语言生活,发现社会语言生活热点和难点,推进应用语言学相关领域的研究。

2. 及时了解社会语言生活舆情动态,促进语言文字工作的宣传和普及,促进政府、专家、公众之间的沟通,为构建和谐语言生活服务。

3. 及时搜集网络舆情热点,研究网络舆情热点形成和发展规律,准确把握网络舆情热点,为政府正确引导网络舆论和制定相关政策提供决策参考。

二、网络舆情研究的现状

关于舆情的概念,学界一般认为有广义和狭义两种。王来华认为,舆情是指在一定的社会空间内,围绕中介性社会事项的发生、发展和变化,作为主体的民众对作为客体的国家管理者产生和持有的社会政治态度^[2]。被学界称为“狭义舆情”。刘毅在此基础上进一步扩大了其外延,提出舆情是由个人以及各种社会群体构成的公众,在一定的历史阶段和社会空间内,对自己关心的或与自身利益紧密相关的各种公共事务所持有的多种情绪、意愿、态度和意见交错的总和。^[3]

网络舆情研究是新兴的社会科学与自然科学交叉的研究领域。国外一些发达国家对网络舆情的研究较早,因为有大量机构和大量专业分析人员的关注,这方面的研究成果相对丰富。在美国,对于互联网舆情研究,特别是与政治生活密切相关的总统选举和国会中期选举的选情分析和研究已达到了很高的水平。

国内的网络舆情研究始于2005年,国家社科基金2005年资助了天津市社会科学院王来华“建立社会舆情汇集和分析机制研究”。从国家自然科学基金和社会科学基金资助的项目看,网络舆情领域的研究项目集中在近两三年内,国家自然科学基金项目主要有:“面向公共安全的网络舆情预警方法研究”(清华大学疏学明主

持)、“基于特征概念网的网上离散文本信息舆情分析研究”(上海交通大学李生红主持)、“网络信息传播的管理机制研究”(北京大学谢新洲主持)等。国家社科基金 2009 年资助了两个重大项目,“突发事件网络舆情预警指标体系研究”(上海交通大学谢耘耕)和“突发事件网络舆情演化模型和仿真系统研究”(解放军信息工程大学邬江兴)。2010 年国家社科基金在政治学、社会学、语言学、新闻学、管理学、图书馆情报与文献学等多个学科资助了近 40 项有关网络舆情研究的项目。目前虽已吸引了相关学科领域的一些专家学者的关注,但总体来看,还很薄弱。从数量上看,网络舆情的文献资料较少(据学者统计,2005—2009 年,CNKI 和维普中文期刊上不到 100 篇),对网络舆情的研究有待进一步丰富和完善。

网络舆情研究的成果主要集中在基础理论研究、关键技术研究 and 应用研究三个方面。

基础理论研究涉及新闻传播学、社会学、教育学、心理学、公共管理学、信息/情报学等多个领域,侧重探讨网络舆情的概念、构成要素、主要特点、主要传播途径,侧重舆情基本机制研究(包括形成模式、变动规律等),舆情汇集和分析机制研究,舆情预警机制研究、舆情监控和引导研究等。论著主要有:王来华主编的《舆情研究概论:理论、方法和现实热点》,刘毅的《网络舆情研究概论》《网络舆情及社会性网络信息传播模式》等。

近几年,随着现代社会复杂程度的进一步提高,世界各国都加速开展了舆情安全的关键技术研究,这些技术研究主要集中在互联网和计算机技术领域。研究范围包括文本数据自动分析、WEB

数据自动分析、舆情感情色彩分析、人工智能检索、自动摘要、决策和复杂性社会网络系统等方面。国外影响力比较大的相关研究成果有:加州大学伯克利分校社会科学计算实验室开展的 SDA 项目,已被广泛使用的 WEB 数据发布和分析系统 NESSTAR,英国科波拉软件公司 2005 年推出的舆情感情色彩分析软件等。

目前国内外网络舆情的相关应用研究主要集中在舆情/情报挖掘、网络反恐和分析民意等方面。国内的应用主要体现在网络舆情系统的开发和专业舆情监测机构的涌现。如方正智思舆情监测分析系统、Autonomy 网络舆情聚成系统、TRS 互联网舆情信息监控系统、GOONIE 互联网舆情监控系统等。人民网舆情监测室(人民日报社网络中心舆情监测室)是国内最早从事互联网舆情监测、研究的专业机构之一,在舆情监测和分析研究领域处于国内领先地位。

2009 年,在教育部语言文字信息管理司的指导和支持下,教育部语言文字应用研究所语言舆情研究中心成立,这是国内最早专门针对语言文字舆情进行研究的机构。中心成立以来,创办了《语言文字舆情与动态》《语言文字舆情月月读》内部刊物,形成数篇监测分析研究报告,在语言文字舆情的监测和分析方面逐渐积累了一些经验。但总体而言,在网络舆情监测理论体系、工作方法、作业流程和应用技术等方面仍需进一步探索和完善。

三、语言舆情监测关键问题与技术方法

对舆情信息的监测与分析必须要从网络新闻报道、相关评论、

论坛等浏览和查找海量的网络信息,从这些信息中提取与事件相关的舆情信息,分析舆情信息的时间与空间分布情况。在海量的信息中,单纯依靠传统的人力搜索已无法有效解决问题。随着互联网技术的不断更新,网络舆情监测和舆情分析有必要通过与之相匹配的科技手段来进行,对网络舆情信息需要进行自动监测与采集。当前社会上的舆情监测系统多由科技企业研发,虽然技术上已相当成熟,但由于缺乏相关的专业知识,所研发的舆情监测系统目前仍无法完全满足专业领域的需求。

考虑到语言文字网络舆情离散性、专业性和多元化的特点,教育部语言文字应用研究所语言舆情研究中心在编辑制作《语言文字舆情月月读》《语言文字舆情与动态》的过程中,不断总结,基于J2EE架构开发了语言文字Web信息搜索与自动聚合系统,主要解决以下几个主要问题:(一)语言文字舆情信息的离散性。考虑到语言舆情发生的特殊性,结合可能形成舆情热点的领域,配置网络爬虫的关键字参数集合,利用网络爬虫对门户、论坛和博客内容进行自动抓取。(二)对异构信息的结构化抽取。爬虫抓取到语言文字话题的网页后利用分析器对其进行结构化信息抽取(可对整个页面进行抽取或仅对正文进行抽取),结合配置文件进行初步数据过滤后形成原始语言文字舆情库。(三)根据不同用户对象进行访问控制。由于舆情信息自身的独特性质,需要根据访问对象不同(如:用户、舆情分析人员或系统管理员)而分配不同权限,本系统采用基于角色的访问控制机制,引入了角色作为用户与权限之间的中介,这个机制将权限同角色联系起来,而对用户赋予相应的角色,用户所能访问的资源权限就是该用户所拥有角色的权限集

合的并集。(四)对聚合信息的管理。为了便于对舆情的深入分析,本系统采用标签(Tag)对舆情信息进行分类,同时提供基于时间和标题关键字的检索。(五)与语言文字门户网站集成。系统整合到中国语言文字门户网中,实现单点登录(SSO),即用户可统一、持续地访问信息,而无需了解也无需关心信息的来源,实现了资源的重用和整合。

自动监测系统如下页图1所示,包括:①舆情监测配置文件:结合可能形成舆情热点的领域,设置关键字,集合为{“书法教育”“方言”“汉字编码”“繁体字”“英语”“简体字”“外语教育”“字母词”“普通话”“汉语热”“孔子学院”“词汇”“字体”“中华诵”“中华赞”“双语教育”“汉字”“词典”“汉字输入法”“语言能力”“通用规范汉字表”“少数民族语言”“汉语教育”“网络词语”“缩略词”“语文教育”“双语教师”“英文”“民族语言文字”“拼音”“语法”“华文”};此外,在配置文件中还设置爬虫抓取页面的时间间隔等。②网络爬虫 Extractor 从门户网站、博客专栏和专业论坛等抓取语言文字话题的网页,然后利用分析器对其结构化,并结合配置文件进行初步数据过滤后形成原始语言文字舆情数据库。③语言文字舆情数据库,其中语言信息数据表主要包括编号、标题、正文、来源、关键字、时间、舆情级别和备注等字段,本数据库开放给舆情分析人员使用,同时作为历史舆情供后续研究参考。④门户舆情频道,包括“前台页面”和“后台审核”。“前台页面”向舆情分析人员自动显示当前最新语言文字方面的报道和评论,每条信息包括标题、来源、来源时间等;“后台审核”对自动采集的报道进行审核,对不适宜的舆情可以删除(如过期、转载或虚假舆情等),以决定是否在前台显示。

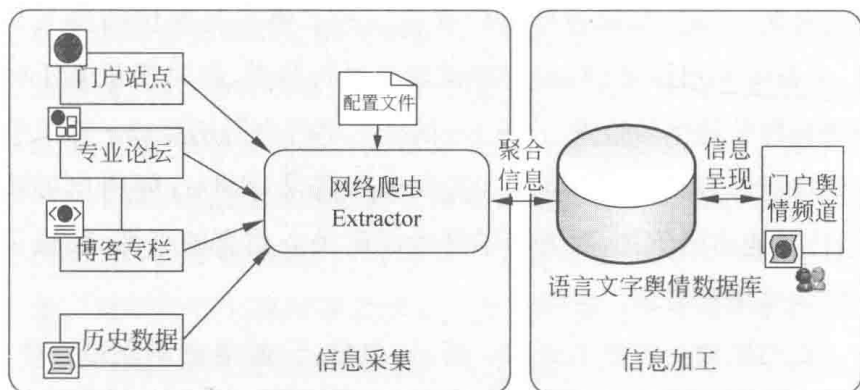


图1 语言文字网络舆情自动监测示意图

当前主要功能主要集中在自动监测方面,后续我们将选取合适的聚类算法通过对文本聚类发现热点话题并进行话题跟踪,通过分析语言文字舆情话题的报道数量和话题持续时间提出舆情事件的热度量化方法。

四、个案分析:关于“广电总局要求媒体限制使用外语缩略词”的舆情分析

2010年4月6日,媒体报道广电总局向央视等媒体发出通知,要求在非外语频道主持人口播、记者采访和字幕中,不能再使用诸如NBA、GDP、WTO、CPI等外语和缩略词,如遇特殊情况必须使用,要在外语或外语缩略词后加中文解释。报道引起了媒体和网民的普遍关注。

舆情研究中心依托语言文字Web信息搜索与自动聚合系统对整个过程的跟踪监测,对其中较有代表性的近120篇相关报道、评论等(不重复累计)进行了分类整理,并搜集一些主要社区

的相关讨论,如天涯社区、网易社区、搜狐社区等,同时参考一些商业门户网站的调查对事件进行分析。

1. 关注度分析。在百度指数输入关键词“屏蔽”,通过用户关注度可以看出,在全国范围内,从4月5日开始,“屏蔽”一词的关注度开始高于均值线,到4月7日达到最高值,至15日以后关注度逐渐低于均值线。从4月6日至18日,除10日外,当日最热新闻都与“屏蔽缩略词”有关。百度指数罗列的4月份包含关键词“屏蔽”的前10个最热点新闻中,与此事件相关的有3个,分别是4月6日、4月7日、4月12日的热点新闻“把CCTV当作外文缩略词”“被屏蔽背后是焦虑与不自信”“屏蔽缩略词应尊重大众习惯”。“央视屏蔽外文缩略词”是4月份用户关注度环比上升最快的相关检索词,上升百分比为100%。从地区分布看,前十位的省份是:广东、北京、江苏、山东、浙江、上海、河南、湖南、福建、湖北,前十位的城市是:北京、上海、深圳、广州、杭州、天津、长沙、南京、郑州、常州。^[4]新浪网调查情况与此基本相似(参见图2)。总体而

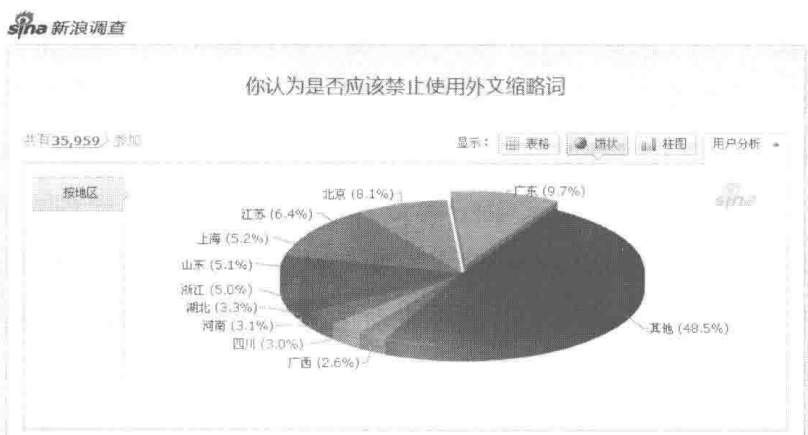


图2 新浪网调查结果图