

大数据时代的政府网络舆情监测系统研究

孙培梁¹ 林枫² 吕志龙³

(1.浙江警官职业学院 浙江杭州 310018;2.浙江理工大学 浙江杭州 310018;3.北京华宇软件有限公司 北京 100084)

摘要:本文围绕大数据时代政府如何加强网络舆情的监测、预警以及分析、引导,如何建设网络舆情监测系统开展研究,对维护政府形象,加强社会公共管理,实现社会的和谐具有十分重要的意义。

关键词:网络舆情 政府 大数据

中图分类号:TP393

文献标识码:A

文章编号:1007-9416(2014)03-0078-02

Abstract: This paper was concerned about how to build a Internet public opinion detecting and analytic system of governments, and how to strengthen the Internet public opinion monitoring, early warning and guidance research. The results of maintain the image of governments, and strengthen public management, has a very important significance harmonious society.

Key Words: Internet public opinion government big data

网络舆情(Internet Public Opinion, IPO)即网络上的社情民意^[1],其概念强调两点:一是新闻事件、社会现象和社会问题主要通过互联网首发或传播;二是表达信念、态度、意见和情绪的公众主要是网民。从上世纪九十年代开始,互联网的飞速发展已经深刻的改变了社会的信息传播渠道;如今普及的移动互联网更是彻底颠覆了传统的媒体和舆论观念,成就了一个“人人都是信息采集者、人人都是信息传播者”的新时代。网络媒体具有进入门槛低、信息规模大、发布与传播迅速、参与群体庞大、实时交互强等综合特点。网络舆论不仅是网民直接表达意见的渠道,更是网民与社会管理者进行互动的绝佳平台,甚至已经开始推动社会在各个领域的变革和进步。

1 大数据时代网络舆情监测的新机遇

社交媒体和传感网络的发展带来数据爆炸式增长催生了大数据(Big Data)^[2]这一新兴概念。大数据所涉及的数据规模如此巨大,以至于几乎无法通过目前的主流软件工具,在合理时间内达到摘

取、管理、处理、并整理成为帮助企业经营决策目的的信息。大数据的数据规模超出传统数据库软件采集、存储、管理和分析等能力的范畴,涉及到多种数据源、多种数据种类和格式。网络舆情信息正是一种大数据,达到甚至超过PB(1000TB)级规模的网络言论事实上代表了网民最真实、最客观的行为特征与心理需求。网络舆情信息的主要来源有:新闻评论、BBS、聊天室、聚合新闻RSS、QQ、MSN、博客、微博和微信等等。传统的依靠人工的方法进行的网络舆情监测在大数据时代显然难以持续。如何时准确地掌握网络舆情动态,提高新形势下舆情信息的分析能力,积极引导社会舆论是各级政府部门面临的严肃课题与严峻挑战。

大数据的基本特征可以从规模、变化频度、种类和价值密度等几个方面进行理解。对于大数据特征的描述集中为5V,即规模化(Volume)、多样化(Variety)、快速化(Velocity)、潜藏价值(Value)以及真实性(Veracity)。

规模化(Volume):聚合在一起供分析的数据规模非常庞大。谷歌执行董事长艾瑞特·施密特曾说,现在全球每两天创造的数据规模等同于从人类文明至2003年间产生的数据量的总和。“大”是相对

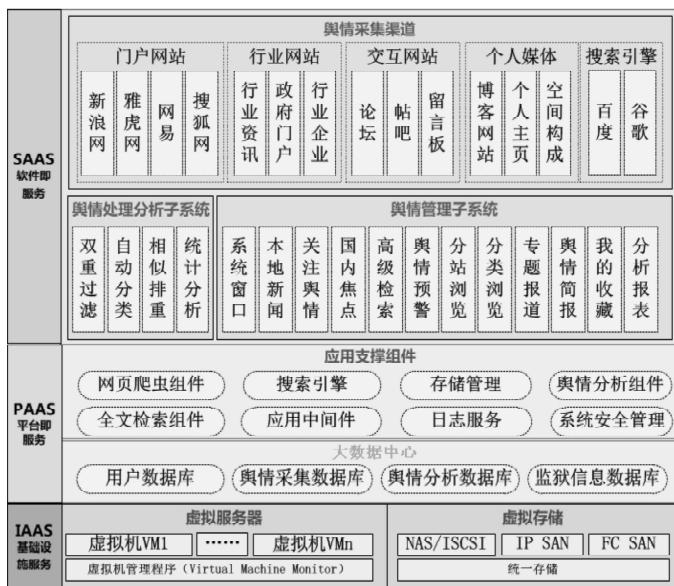


图1 政府网络舆情监测系统“云平台”总体架构图

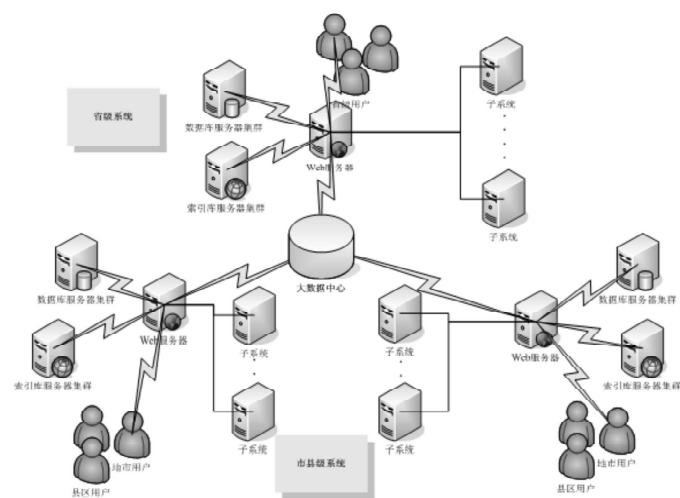


图2 政府网络舆情监测系统的部署

* 基金项目:本文为2013年浙江省人力资源和社会保障科学课题“大数据时代的政府网络舆情应对研究——以司法系统为例”(项目编号:R2013A008)的阶段性研究成果。

作者简介:孙培梁(1977-),男,浙江绍兴人,浙江警官职业学院副教授,研究方向:司法舆情、监狱物联网。

而言的概念,对于搜索引擎,EB(1000PB)属于比较大的规模,但是对于各类数据库或数据分析软件而言,其规模量级会有比较大的差别。

多样化(Variety):数据形态多样,从生成类型上分为交易数据、交互数据、传感数据;从数据来源上分为社交媒体、传感器数据、系统数据;从数据格式上分为文本、图片、音频、视频、光谱等;从数据关系上分为结构化、非结构化、半结构化数据;从数据所有者分为公司数据、政府数据、社会数据等。

快速化(Velocity):一方面是数据的增长速度快,另一方面是对数据访问、处理、交付等速度的要求快。美国的马丁·希尔伯特说,数字数据储量每3年就会翻1倍。人类存储信息的速度比世界经济增长速度快4倍。

潜藏价值(Value):尽管我们拥有大量数据,但是发挥价值的仅是其中非常小的部分。大数据背后潜藏的价值非常巨大。美国社交网站Facebook有10亿用户,网站对这些用户信息进行分析后,广告商可根据结果精准投放广告。对广告商而言,10亿用户的数据价值上千亿美元。2012年,运用大数据的世界贸易额已达60亿美元。2016年,这个数字预计将达200亿美元。

真实性(Veracity):一方面,对于虚拟网络环境下如此大量的数据需要采取措施确保其真实性,客观性,这是大数据技术与业务发展的迫切需求;另一方面,通过大数据的分析,真实地还原和预测事物的本来面目或者是未来的发展趋势也是大数据发展的关键问题。

2 政府网络舆情监测系统的建设需求与目标

政府网络舆情监测系统建设的实际需求有:

- (1)如何透过互联网了解民情,获取最真实的民众诉求?
- (2)如何准确地收集‘政府各部门最需要的’舆情信息?
- (3)如何能随时知道‘与政府各部门相关舆情信息’的发生,以及如何全面追踪?
- (4)如何应对突发事件和舆情危机?
- (5)如何防止有害舆情信息的泛滥传播和舆情失控?
- (6)如何判断舆情信息的未来走势?
- (7)如何在第一时间为政府各部门领导决策提供信息支持,为危机化解争取时间?
- (8)如何建立舆情评估机制?

因此政府网络舆情监测系统的建设的目标是:

借助大数据支撑技术,通过对主流媒体、门户网站、资讯平台、知名论坛、搜索引擎、博客、贴吧等网络载体的监测,全面掌控互联网上与政府各部门相关的人、地、物、事、组织,不漏掉有价值的舆情信息。借助大数据支撑技术,对已经获取到的海量网络信息中与政府各部门相关的舆情信息进行动态分类采集、智能化分析,及时发现隐藏热点舆情、突发事件和重大公共事件信息。借助大数据支撑技术,实现第一时间掌握涉及政府各部门相关的网络舆情最新动态。

3 政府网络舆情监测系统“云平台”构建

3.1 舆情系统总体架构

政府网络舆情监测系统“云平台”总体架构如图1所示。

政府网络舆情监测系统“云平台”通过虚拟化技术^[3],在物理服务器的基础上搭建若干虚拟机,部署相应的应用服务器、数据库服务器,并基于NAS^[4]、IP SAN^[5]、FC SAN^[6]等技术构建虚拟化存储体系,在云端存储所有的应用数据。通过云平台管理套件实现虚拟服务器、存储资源的定制化、可计数、动态分配和使用,并在“云平

台”的基础上搭建网络舆情监测系统应用。

3.1.1 基础资源服务(IAAS)

在电子政务专网、服务器、存储设备等IT基础设施的基础上构建政府网络舆情监测系统“云平台”的基础资源服务,包括:虚拟服务器、虚拟存储等。

虚拟服务器:在物理服务器上安装虚拟机管理程序(Virtual Machine Monitor),虚拟机管理程序可根据用户实际需求(CPU主频、内存大小、硬盘容量等)构建相应的虚拟服务器作为部署应用系统的应用服务器和数据库服务器。

虚拟存储:通过成熟的NAS、IP SAN、FC SAN技术构建集群的云端存储体系,为各个虚拟应用服务器和数据库服务器提供存储。

3.1.2 平台即服务(PAAS)

通过“云平台”管理程序对虚拟机的创建、负载均衡、回收、备份进行统一的管理,并提供相应的技术组件为“云平台”上的应用提供支撑。政府网络舆情监测系统通过PAAS层向用户提供舆情监测相关的功能服务。PAAS层是为上层服务提供技术支持的组件和数据库服务,支持组件包括网页爬虫组件、搜索引擎、存储管理、舆情分析组件、全文检索组件、应用中间件、日志服务、系统安全管理等组件。

大数据中心包括用户数据库、舆情采集数据库、舆情分析数据库、监狱信息数据库等等,基于以上各类数据,包括所有格式的办公文档、文本、图片、XML、HTML、各类报表、图像和音视频信息等等结构化、非结构化、半结构化数据,由于结构和来源不同,系统提供了不同的访问方式和相关协议标准。

3.1.3 软件即服务(SAAS)

软件即服务(SAAS)是在基础资源服务(IAAS)和平台服务(PAAS)的架构基础上,为政府用户提供提供网络舆情应用服务,主要通过舆情采集、舆情处理分析、舆情管理三大子系统提供。

3.2 舆情系统部署

政府网络舆情监测系统可以按省、市县等层级实施部署。通过电子政务专网,将政府各部门有机连接,共建舆情大数据与预警中心,整个网络拓扑架构如图2所示。省级系统和市县级系统分别独立,上级主管部门可以访问下级采集的数据,如果达成协议,同级各部门也可以互相共享采集回来的数据。

4 结语

网络舆情已成为影响社会持续有序发展、维护社会和谐稳定的重要因素,政府各部门有必要高度重视网络舆情监测系统的建设,积极应对大数据环境下互联网信息内容的管理、信息基础设施的保护,通过网络舆情的监管和引导,依法治理网络空间,维护公民合法权益。

参考文献

- [1]法制网舆情监测中心.政法舆情危机应对实务手册[M].北京:经济管理出版社,2013.
- [2]维克托·迈尔·舍恩伯格.大数据时代[M].浙江:浙江人民出版社,2012.
- [3]虚拟化技术.<http://baike.baidu.com/view/13605.htm>.
- [4]NAS.<http://baike.baidu.com/subview/56335/7913805.htm>.
- [5]IP SAN.<http://baike.baidu.com/view/1810605.htm>.
- [6]FC SAN.<http://baike.baidu.com/view/1810601.htm>.