

档案学视角下网络舆情大数据的采集、研判和归档研究

顾 琨

(黄河水利委员会黄河中心医院 河南 郑州 450004)

【摘要】随着互联网行业的兴起,网络舆情大数据资料就如同纸质资料一般,能够进行保存存档。它能够长久的将文字以及图片保存下来,能够更加清晰的反映出以往事迹,因此它具有长期保存的意义。但是随着社会的发展,网络舆情大数据就越来越复杂,对有价值的网络舆情大数据的采集、研判以及归档提出了更高的新要求,为此,我们需要对网络舆情大数据的研究从档案学视角下进行。本文主要分析了司法鉴定方面网络舆情大数据的采集、研判和归档的研究。

【关键词】舆论大数据;采集;研判;归档研究

【DOI】10.12252/j.issn.2096-6288.2020.06.180

引言

在大数据的发展背景下,网络数据开始慢慢转变为一种基础性的资源,大家越来越关心对网络数据的利用、采集与管理。网络舆情反映出大家的意见,影响了社会各个方面,而由于现在社会舆论越发厉害,许多案件最后的司法裁决也被社会舆论所左右。网民们积极主动在网上发言,虽反映民意,但是可能与司法相违背,在网络大背景下应积极处理理论与社会各种现象的冲突,积极回应网民对社会公平正义的要求。

一、网络舆情大数据的采集

网络舆情是从各类事件的网络报道中入手,民众对所报道事件形成自己独特的见解。网络舆情和传统舆情不同,因借助的媒体设备或传播媒介是网络,因此网络舆情具有虚拟、传播范围广、信息多元且分散的特点。网络舆情关注的更多是媒体和互联网争论的焦点,因此依托的平台具有极高的瞩目性和公众影响力,网络舆情体现出针对性的特点。在对各项事务的影响因素进行分析的过程中发现网络舆论会影响到公民的个人判断,还可能影响到事情的发展态势。网络舆情还具有群体性的特点。关注某些信息的群体交换意见,以共有的情感分析事态走向。网络舆情周而复始,循环往复,从事态的苗头开始,逐渐扩散、发展,中途达到网络舆情的发展高峰,再回落下降,一般以年或者月为单位进行计算和收集。网络舆情大数据的采集大概分为数据的采集、数据的预处理以及对数据的存储三个步骤。

1. 数据的采集

在数据的采集时,运用一些高新技术对舆情数据进行自动化批量的采集,采集的时候,要保证信息完整、准确,并且将繁杂的数据进行分析简化,变成可以让人更加容易理解的数据。数据采集可以借助一定的技术,例如大数据存储技术,实现对采集信息的高吞吐量的收集和分析,并将收集到的信息应用到开发分布程序,增加分布式的底层细节。利用微信、微博等对网络舆情数据进行实时监控,将数据智能连接,建立智能采集系统,分析数据间的关联性。

2. 数据的处理

在数据的预处理中,要对原始的数据进行二次过滤,删除繁枝末节、不符合人之常情以及违背法律的数据,从此确保采集到的大数据客观真实。像司法鉴定的大数据采集时的预处理,要将违背法律的、无理取闹的舆论剔除,保留有利于司法鉴定的舆论,最后处理的结果要使民意和法律体现的淋漓尽致。若想预先理解人们无法得知、不了解、隐藏的但又对舆情分析十分有用的信息,可以采用大数据挖掘技术,利用此技术将大量、随机分布、不完全处理的舆论搜集出来,方便后续操作。

3. 数据的存储

将数据收集和预处理完成后,要对数据进行储存,存储一般情况下将大数据保存到多个不同的远端服务器,这样起到备份的作用,防止存储设备出现故障和防止存储信息丢失^[1]。

二、网络舆情大数据的研判

在如今网络舆情与司法审判价值取向相矛盾的社会,对网络舆情大数据的科学研判至关重要。因此对网络舆情大数据进行研究判断时,要利用现在的信息研究手段,对数据分别进行定量、定向和定点的分析,只有这样,才能使大数据更加清晰的反映出在司法鉴定方面的民意,有利于司法鉴定,有利于社会积极的发展。

大数据的应用处理和辅助分析系统可以为工作人员提供可靠的数据分析,调整案件分析的方向。司法舆情案件由大数据进行评估,在不同时期将工作重点转移,第一时间分析处理措施。每天均会出现各种各样的网络舆情分析案件,为了优化大数据分析处理的流程,尽量减少舆情回应的平均时间,优化审批处理流程,采用简化、规定的流程对各项热点司法案件进行追踪,将计算处理流程缩短到分钟。以下是对数据处理分析的具体内容分析:

1. 对数据进行定量分析

对大数据进行定量分析的意思就是说要大数据进行量化处理,如今网络信息技术发展迅速,要想对网络舆情大数据进行正确的判断研究,定量分析必不可少。一个完整的舆情信息包括诸多方面,因此在量化分析的时候,要重视分析直接或间接与该司法事件相关的各个帖子的发帖量和跟帖量。加使用大数据等新兴技术手段的运用和处理,不断鉴别与司法案件有关数据的解读、分析,对数据进行定量的加工分析处理,增加数据的基值,为科学判断提供理论依据。

2. 对数据进行定向分析

由于现代网络的发达,网络舆情的出现是一个迅速的过程。首先网络上出现舆情信息,网民就通过点赞、发帖、跟帖来表达自己的意见,但是网民在表达意见的时候,他们会出现意见分化^[2]。因此各个司法舆情的信息非常多,数据的研究方向也很广,因为人力物力有限,不可能逐个方面进行分析,因此要判断该网络舆情的某个意见是群体的意见还是某一特定群体的意见,然后选好研究方向,只有对舆情大数据进行定向分析,才能使司法判断更加公平公正且符合民意。根据社交媒体发布的各项信息,将数据进行深入的加工处理,将信息进行精确定位,记录某位公民登录某个软件的事件、地理位置、关注案件的类型等信息,将这些内容进行精确定位,实现定向、高效、准确地监管和分析,为司法舆情分析提供优势。

3. 对数据进行定点分析

从整体来看,网络舆情反映的民众的意见,而网络舆情最后的发展结果也是被大多数网民推动的,网络舆情由量变转变为了质变,但是这些意见的显现都是由一些关键性的网民推动,这些网民是我们对突发群体事件进行沟通和处置十分关键。

三、网络舆情大数据的归档研究

1. 对网络舆情大数据进行归档的原则

在归档时注意前瞻性、实用性。采用大数据分析平台将归档的业务管理和系统操作效率提升,为互联网的便捷使用和分析提供接口。其次要遵循灵活性和扩展性的原则,将归档系统进行延展。除此之外,还要对文档进行数据保护,确保文档和收集的数据的准确性和安全性。下面是对网络舆情大数据归档原则的具体分析:第一要对有价值文档和无价值文档进行区分,对文档的真伪性进行鉴定,判断该文档是否有保存的价值,因为文件内容多样复杂所以要科学客观的进行判断^[3]。第二要拟定网络文档归档的范围和归档的时间。第三要选择合适的收集方法进行收集,收集时要明确各个收集机构的职责,使他们各司其职,最后进行归档处理。

2. 对网络舆情大数据进行归档的方法

大数据舆情管理分析系统的本身是一个辅助管理的数据运行和分析系统,将各种数据进行归档后不断分析。其中数据的归档输入包括从各类渠道中搜索有用的信息,例如微信、微博、各种新闻网站首页、常用搜索网站的信息。数据的输出包括对案件的分析处理,对归档后结果的调查、对舆情的发布。数据分析包括对网络舆情的分析、对司法案件的分析。最后数据管理服务是归档后对各项任务、获得的各种文案、设计的各项流程的管理。对获得数据进行归档处理时,主要分为逻辑归档和网络实时归档。逻辑归档是利用网络信息的优势,不改变大数据原来的存储方式和位置,将大数据信息移交到相关归档部门。逻辑归档简化了归档流程,将计算机的共享优势发挥到了最大,但是由于网络技术发展不完善,所以逻辑归档要谨慎。网络实时归档是利用计算机网络对其进行管理,保证了归档信息的原始性、可靠性和完整性。

网络舆情分析的过程对精细程度要求较高,加上相应工作人员的缺乏,使得我们在对网络舆情案件分析时缺少精准分析。相关管理分析人员应根据收集获得的数据进行有针对性的分析,形成共同的目标,从档案学的视角完善管理体系,分工合作、简化归档流程,各部门联合将数据分析到位,为司法案件的处理提供依据。

结束语

随着网络技术日益蓬勃发展,网络舆情也在不断的发生变化,网络舆情大数据我们要通过档案学视角进行有效的收集、研判和归档,只有对大数据进行按步采集,定向、定量、定点研究判断,注意归档的原则和方法,才能使司法鉴定在前行的道路上越走越远。

参考文献

[1]冉朝霞.档案学视角下网络舆情大数据的采集、研判和归档研究[J].档案管理,2018,(1):57-58.

[2]范胜华.大数据时代网络舆情监管机制研究[D].华中师范大学,2019.

[3]张思龙,王兰成,姜国哲,高宾.大数据时代网络舆情研究的前沿演进和发展趋势[J].中华医学图书情报杂志,2019,28(01):50-55.

作者简介:

顾琨(1985-),女,汉族,河南郑州人,法律硕士,任职于黄河水利委员会黄河中心医院河南唯实司法鉴定中心,河南,郑州,450004,研究方向:档案管理,司法鉴定机构管理