

大数据背景下网络信息挖掘与情报收集

叶胜万

(武警湖北省总队综合信息保障中心,湖北 武汉 430000)

摘要:如今,随着我国网络电子信息、互联网通信多媒体等新一代信息网络技术的快速进步和蓬勃发展,网络上的海量信息日新月异,大数据在我国乃至全球范围内迅速发展,掀起了一股对网络大数据进行信息收集和分析的狂潮,成为当代全球和人类社会主义市场经济社会的重要信息技术力量和生产力。当前的技术存在问题:网络上巨大的信息安全,巨大环境下的巨大数据分析,信息收集与分析技术,大数据收集,综合分析与信息收集,数据分析技术不仅仅是对这些海量信息的综合收集、分析或收集,也绝对应该只是对这些巨大信息的综合收集、分析和逻辑优化,从而整合这些巨大信息背后潜在的巨大技术价值。大数据应用技术将有机会给整个人类经济社会发展带来新的经济改革和发展方向,促进整个人类经济社会发展的文明和谐发展。主要探索分析网络应用环境下大数据技术发展的基本规律,分析网络应用环境下大数据技术的各种特点,针对网络大数据信息资源的有效管理和应用创新途径,进一步研究其他大数据的信息采集、分析和处理应用技术。

关键词:网络环境;大数据;采集和处理

网络大数据主要是指远远超过原有数据系统实时采集和传输处理信息能力的网络数据,是当前网络应用环境中常见的数据处理信息。由于大数据对网络信息采集规模和数据传输处理能力有很高的技术要求,又由于与原始数据系统不完全匹配,大数据中的信息不会被数据系统及时分析、整理和收集处理。大数据在当今信息时代对技术发展和进步起着重要的推动作用,能够有效促进新信息生产力的快速发展。因此,在电子信息应用技术不断变化和变革、发展速度不断扩大和加快的电子信息时代,为了及时了解和充分认识可能隐藏的重要技术信息及其应用价值和功能,促进电子信息生产力的创新发展,相关技术专业的信息技术人员应及时采取行动,采用有效的本地方法,及时收集和分析信息^[1]。

1 网络环境下大数据的发展

大数据的新创业时代即将开始。其中一家定义明确的创业公司是美国的咨询服务公司麦肯锡,这家企业咨询管理公司在美国社会发展研究部的市场调查报告明确指出,大数据时代的信息在当今社会各行各业得到了广泛的整合和应用,已经成为当今社会各领域重要的核心生产力。未来,随着各种海量大数据的广泛使用,再次预示着新一代信息和工业生产力快速增长的新时代的到来。大数据未来的发展趋势是未来人类社会持续进步、健康发展的重要标志,正预示着未来人类社会朝着更加文明、科技化、信息化的方向持续进步、健康发展。在整个现代信息社会中,大数据逐渐成为现代计算机和网络媒体信息技术研究的热点研究内容。

2 大数据背景下的网络信息挖掘

政府部门对行政网络宣传信息的及时收集、分析和处理,主要是借助移动互联网技术,对政府网络上传的这些信息来源及时进行收集、分析和整理,并有效利用这些网络信息源,为今后的网络工作提供良好的网络环境。如果研究这两个方面。发展效果做得好,可以有效的来提高其他政府部门的网络工作效率,进一步可以实现其他政府部门或专门对政府网络宣传舆情的有效治理。一般说的情况下该分析部分主要包括对于在网络上的内容、结构和内容用法的及时挖掘。首先是在网络上的内容结构具有网络复杂性的基本特点,比如大量文字、图片等等,对于与网络相关内容的及时结构挖掘对于可以快速获得有关社会动态的第一时间资讯,具有重要指导意义。其次,

网络内容结构挖掘需要同时实现对超链接、内部结构和URL三个基础方面的及时挖掘;第三,网络内容用法的及时挖掘则主要目的是为了帮助网络分析师找出每个用户的网络访问量和偏好,为之后的网络决策工作提供参考依据^[2]。

3 大数据专业

3.1 以实时性特点划分,分为实时数据分析和离线数据分析

实时数据分析主要应用的专业领域主要指的是各类B2C分析产品,包括企业金融、互联网和其他移动互联领域内的各类B2C分析产品。这类分析产品往往通常会根据要求在短期限内一段时间内每周反馈超过上亿次的用户信息,以此被用来作为达到不断提高企业用户操作体验的主要目的。实时数据离线用户数据处理信息统计分析对实时用户反馈处理数据信息统计体系的处理技术要求比离线实时用户离线数据分析的技术需求高度要高,而当前对于重大事件数据采集和时间转移信息处理的缺乏深入理论研究也主要原因是其研究针对的主要问题之一是如何利用离线实时用户离线数据分析处理信息。对于大多数对实时用户离线反馈处理信息统计时间转移处理能力要求不高的大型企业应用,如通过离线实时机器学习深度学习、推荐数据搜索分析引擎离线数据转移计算、离线实时用户反馈统计数据转移分析等,可以优先考虑选择采用离线实时进行离线数据分析的转移处理方式,通过根据用户反馈采集的大量实时离线数据将其处理信息体系转移到并处理到能为用户提供专用的离线数据分析转移处理服务平台^[3]。

3.2 以数据量分析,分为内存级别、海量级别、BI级别

群体海量类型级别一般上就是指需要广义泛指不一定属于具有超过一个用户群体的企业产品自身内存最大数据容量的数值。都是其需要是以特定统计数据量级和类型的作为主要分析基础的数据来直接进行统计数据海量分析,同时也是其需要能够满足的要求是对于大数据的一个基本概念。所以,在具体的实际网络应用中我们常常需要对数据类型是否可以及时进行正确的分类或者进行比较,以便于我们采取适当的网络信息数据采集及相关数据处理技术分析方法。提高海量网络数据采集级别的大趋势海量网络数据采集资源和大流量数据采集技术是当前我国网络通信发展的一个主要发展趋势,以便于确保海量信息数据特征的有效准确分类和易采集和及时处理。

4 网络大数据采集的基本流程与关键技术

4.1 网络大数据采集的基本流程

技术管理流程系统是网络大数据采集的一个基本整体,其技术流程框架主要可以分为一个包括六大基本技术流程板块:主要包括分析网站优化目录管理页面、链接管理数据采集抽取、链接管理数据抽取过滤。然后自动直接提取本站网页所收录需要的网页内容值。写入新的网站数据库,放置已分类抓取的网络链接,对分站站长的网站属性链接源地址进行最终分类抓取;最后,集中获取具有新网页属性链接源地址的网站的网页内容,并提取新网站所需的具有网页属性的网站内容的提取值。最后一次将未分类抽取的网页内容抽取值之和写入新网站数据库。要对海量大数据的信息进行全面的分析和处理,就必须完整地建立一个组织良好的大数据处理技术的基础工作流程和信息分析方法,这样才能收集各种信息源,对我们有序处理的海量大数据信息进行全面的分析和处理。此外,在分析和处理这些海量数据时,应用从海量数据中收集的信息的一些基本特征和特性,进一步充分挖掘信息源,对分析数据进行综合处理,以保证其处理的准确性。

4.2 数据采集的关键技术

数据采集的一项关键技术是布隆滤波。链接布隆过滤器是该功能技术的一个主要核心技术功能,即可以自动判断当前网站链接中的内容数据是否在搜索引擎已经自动抓取的当前网站链接中。当引擎需要手动采集当前网站链接的大数据时,可以通过手动添加布隆过滤器的功能模块,自动采集并过滤当前网站链接。关于如何使用布隆布隆滤波器的基本设计思路是:左右。如果节点集中的元素需要手动添加新的节点元素,则可以通过手动将针对x个的新元素集哈希映射函数添加到组值来实现自动捕获新添加的元素,并且可以将x个元素哈希映射函数应用于每个节点的每组收集数据,并且将这些哈希映射节点数据的每组的输入值设置为1。在整个元素检索过程中,我们只需要仔细观察这些点的个数是否必须为1,基本上就可以准确判断元素集中是否有这个新元素。

5 新时代背景下的新挑战

5.1 数据处理能力方面

目前,在大数据发展背景下,虽然大部分地方公安情报部门都成立了自己的省级情报数据合成技术研判机构,但情报数据合成、处理、分析和应用能力水平与传统大数据融合的实际应用需求和实际情况并不完全同步,缺乏集约、高效、智能、便捷的情报数据处理应用工具,对于新信息技术的融合应用严重不足,新技术支撑平台项目建设进度滞后。如何进一步提高情报数据处理和应用能力,是当前地方公安机关情报部门需要重点关注的重要技术问题。

5.2 移动端的整合应用方面

随着当前移动智能手机、平板笔记本电脑等移动网络终端使用成本大幅下降,设备性能和用户操作体验显著提升,在移动商家逐利和移动用户消费需求双增长的刺激下,网络跨应用不断向当前大、中、小规模移动网络终端迁移。如何更好地挖掘当前移动网络终端上的网络交叉信息,如何对当前传统网络智能信息采集处理方式和移动网络信息应用在大、中、小规

模双向迁移的大趋势下逐渐形成的网络交叉信息领域的应用进行深度整合和再整合,是当前网络环境下需要持续认真思考的新技术课题。

5.3 分析技术方面

由于大数据所具有的“大”、“快”、“杂”等基本特征,充分利用传统“云信息数据分析”的现代情报分析研究技术和研究方法,显然已经达到了落后于当前社会环境的情报现实的发展要求。要积极探索创新发展,将其与大数据时代背景下的情报分析研究技术平台相融合。可以建立综合性的情报分析研究技术平台,创新的情报分析技术方法和研究模式,与大数据形成更多的相互关系。

6 大数据背景下网络信息挖掘的举措

6.1 搜索引擎

对舆情信息的分析和收集我们需要及时对这些网络舆情信息源进行分析和搜索。搜索网站引擎不仅可以为您实现更有效、更有针对性的实时查询,还可以及时整理收集这些网络信息源,甚至在必要时及时确认这些信息的数据源,对于鼓励政府部门自行进行信息分析和收集网络舆情数据源具有重要意义。

6.2 信息查询

在互联网使用条件下,很多信息操作也需要对通过互联网技术上传的信息进行实时操作,如输入手机号码等。政府部门在获得所需信息时,也可以通过各种互联网技术及时直接查询信息。

6.3 访问信息

访问信息主要查询功能我们指的也其实就是帮助相关部门发表历史信息工作人员的及时快速浏览全部所有历史,对这部分及时浏览所有历史的及时分析探寻不仅真的可以及时快速实现对帮助政府及其相关部门发表信息人员的喜好等相关历史数据信息的及时分析获取,在帮助政府部门进行有关网络言论舆情的历史信息处理和数据分析的相关工作开展过程中它还真的可以及时快速实现帮助政府及其相关部门工作人员的及时浏览全部历史,获得一些政府必备的相关历史数据信息。

7 结语

当前网络经济环境下基于大数据的整体信息展现类型必然是丰富多样的,需要实时采集和分析处理的信息数据量庞大,对不同信息类型数据整体展现的精度要求也比较高。大规模数据技术是能够促进我国社会主义生产力新方式发展的重要关键因素,是能够引起推动人类社会变革的重要组成条件。所以,在当前网络发展环境中就需要特别抓紧对网络大数据采集和分析处理相关技术的深入研究,以不断提高网络大数据的采集高效性和技术可用性,促进我国社会经济生产力的健康发展。

参考文献

- [1] 王文斌,张磊.大数据背景下网络信息挖掘与情报收集[J].湖北警官学院学报,2016,29(3):60-64.
- [2] 徐颖.大数据背景下的网络信息安全控制机制与评价体系[J].IT经理世界,2020,23(10):194,196.
- [3] 邹士利.大数据背景下网络信息安全问题与对策[J].商品与质量,2019(36):200.