

浪潮舆情云监测系统 技术白皮书

浪潮软件集团有限公司

云服务事业部

2014 年 11 月

目录

第一章 浪潮舆情云监测系统介绍	3
1.1 概述	3
1.1.1 舆情监测背景.....	3
1.1.2 舆情监测工作面临的问题.....	4
1.1.3 产品简介	5
1.1.4 产品主要功能.....	5
1.1.5 浪潮舆情云监测系统常用术语.....	6
第二章 产品设计	7
2.1 产品总体设计	7
2.2 产品设计原则	8
2.3 浪潮舆情云监测系统架构.....	9
2.3.1 信息采集	9
2.3.2 信息处理	14
2.3.3 信息分析	18
2.3.4 信息展示	23
2.3.5 其他	25
第三章 技术优势	27
3.1 技术成熟性	27
3.2 业务需求可靠性	28

第一章 浪潮舆情云监测系统介绍

1.1 概述

1.1.1 舆情监测背景

目前,中国有大约 3.83 亿网民,其中包括 2.26 亿网民看网络新闻;1 亿网民访问 BBS;1.82 亿网民开博客;1.55 亿网民使用手机上网,2.4 亿使用即时通讯工具。网民成为网络舆论的主要推动者。

我国近年来经济体制加速改革,利益格局深刻调整,正处于社会转型、矛盾加剧的特殊时期,公共危机事件由偶发变为频发。而在网络中,信息传播与意见交互空前迅捷,网络舆论的表达诉求也日益多元化,庞大的网民群体的参与使得公共危机事件在短时间内迅速扩散,增强了相关事件的变化因素。

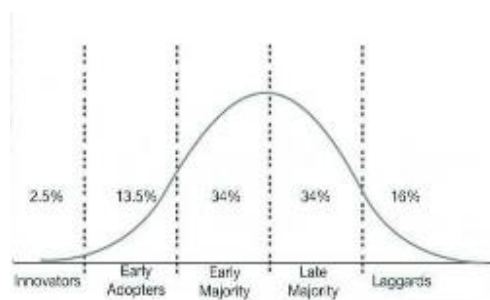
同时,我国国家领导人高度重视网民意见。自 2003 年起,舆情监测成为各级党政部门的一项重要工作,开始进入制度化、规范化发展阶段。此后,我国陆续建立了从中央到地方的舆情监测体系。2006 年 11 月,《中共中央关于构建社会主义和谐社会若干重大问题的决定》指出,“健全社会舆情汇集和分析机制,完善矛盾纠纷排查调处工作制度。”

要建立由中央至地方、涵盖不同部门的全方位的综合舆情监测体系。各相关部门不仅要做到从大处着眼,关注全国范围内有普遍影响力的议题,更要有针对性地对自身管理的领域实施舆情监测。同时,要与其他部门建立联系,共享资源,连点成线,结线成面,形成全领域、多角度的舆情监测体系。可以通过设立网络

监察员等方式,随时了解网络动态,随时掌握并处理网民共同关注的事件和议题。还可以建立网络舆情分析中心,提高对于网络舆情的处理技术,实现利用信息技术对网络群体事件进行精确的分析与预警。

1.1.2 舆情监测工作面临的问题

网络舆情形成迅速,并且其呈现的影响力日益增加,已经渗入到政府政策发布、企业活动等社会的各个方面,互联网已经成为思想文化信息的集散地和社会舆论的放大器。网络舆情具有传播范围广、速度快、影响大的特点,具有直接性、突发性和偏差性的特点,具有易变性和非稳定性的特点。舆情带有倾向性和价值选择双重特征,具有很大的情绪化倾向,具有一定的凝聚力和煽动性。



网络舆情传播规律

传统的舆情监测手段具有舆情收集困难、获取信息不及时、获取信息不准确等局限性,这就导致了各级政府部门不能及时掌握有效的民众舆情,增加了政府对社会管理的难度;增加了社会不安定因素;降低社会公众对政府的信息度;危害政府公共管理秩序。

面对鱼龙混杂的网络社会,各级政府部门如果能够及时掌握舆论的动向,快速分析舆论趋势,便可以积极引导舆论走向。可以充分利用网络舆论分析工具,加强行业自律,改变工作作风、积极有效地宣传政策方针,维护社会稳定,真正

做到关注民生、重视民生、保障民生、改善民生。放大正面声音，宣传积极向上的事物；缩小负面声音，及时处理突发事件造成的不良影响，减小事件的社会危害性。

如何及时掌握舆论动态，快速分析舆论趋势，把握舆论走向，传统的人工调查方式已经无法满足当今网络时代的要求，利用现代网络技术的海量数据搜索、分析、处理能力解决对舆论的监测分析已经成为必然趋势。

1.1.3 产品简介

浪潮舆情云监测系统，是利用爬虫及搜索引擎技术对互联网的标题、网页内容、论坛、BBS、贴吧、微博、作者、时间、出处等各个方面及多个维度进行数据采集，然后通过基于自主知识产权的专利算法进行深度的数据挖掘以及智能化的数据分析，从而生成 Word、Web、PDF 等各种形式的报表，称之为舆情简报。

舆情简报可为企业用户提供各类信息，包括但不限于政策、行业动向、友商动态、负面新闻，从而为企业用户提供各类决策依据，有助于企业用户加强品牌建设。同时舆情简报可为政府用户提供舆论热点分析、民意搜索与调查、负面舆论的搜集、消除及引导等功能。

1.1.4 产品主要功能

浪潮舆情云监测系统利用国内领先的爬虫技术，实时监测新闻、论坛、博客、微博、平媒等网络舆情信息，并对敏感信息及时预警。系统利用中文分词技术、自然语言处理技术、中文信息处理技术，对信息进行垃圾过滤、去重、相似性聚类、情感分析、提取摘要、自动聚类等处理，自动生成个性化的舆情监测报告。

浪潮舆情云监测系统主要包括舆情信息采集, 舆情信息处理, 舆情图表分析, 多维度微博监测, 舆情危机预警, 舆情统计报告等功能。

1.1.5 浪潮舆情云监测系统常用术语

➤ 关键词

指输入搜索框中的文字, 是用于搜索新闻、信息的主题内容。

➤ 报表

指通过关键词所抓取在互联网上发布的新闻列表和分析图表的信息汇总。

➤ 数据

指通过关键词进行系统抓取后对已有新闻信息并进行不同维度存储的数据。

➤ 属性

系统对数据进行正面/负面/中立进行判断。

➤ 监测

指本系统连接到远程管理对象上进行数据采集并对所采集数据进行状态判断的一个流程

➤ 预警

预警方式包括邮件、手机短消息、MSN、GTALK 等方式。

➤ 舆情分析

帮助用户进行分析的图形和数据包括舆情走势、媒体覆盖、来源网站、属性对比表等, 表现形式包括曲线图、柱状图、面积图、饼图、正负图。

第二章 产品设计

2.1 产品总体设计

针对基于互联网的各类舆论信息，我们认为应在统一的信息安全体系及应用系统管理体系下进行规划与设计，在一个统一的基础框架基础上建设各应用系统。

系统利用搜索引擎技术、自然语言处理技术、知识管理方法，通过对互联网海量信息自动获取、提取、分类、聚类、主题监测、专题聚焦，实现用户对网络舆情监测和热点事件专题追踪等需求，形成舆情决策库、舆情简报等分析结果，为客户全面掌握舆情动态，做出正确舆论引导提供分析依据。

浪潮舆情云监测系统具有舆情采集、舆情分析、舆情预警、舆情搜索、舆情简报、统计分析等核心功能。浪潮舆情云监测系统实现了 7*24 小时跟踪、监控、采集目标网站、论坛、微博、QQ 等即时通讯工具和博客的动态舆情信息，及时准确地为用户收集整理关注领域内的热点新闻、突发事件、人物活动等舆情资料；通过敏感话题热度分析，为用户从海量舆情信息中分拣出最新重大舆情事件，让用户实时掌控热点舆情及行业动态，对突发事件和负面舆情作出及时预警；通过跟踪监控舆论报道的发展，分析舆情信息的传播途径，追溯危机舆情的根源，为及时防范处理负面舆情泛滥和扩散、积极引导事态发展的趋势、协助用户将危机消除在萌芽状态，提供高价值的参考决策及危机处理服务。

2.2 产品设计原则

根据国家相关监管部门的要求，本产品的规划和设计参考以下的原则：

➤ 合规性原则

产品的整体安全防护设计必须符合国家、地方的相关标准和要求。严格按照标准进行设计。

➤ 易操作性原则

浪潮舆情云监测系统具备完善的功能，操作简单、直观、维护管理方便。充分采用先进技术，尤其是在信息采集、分析、管理技术等方面，采用领先的智能化技术。在管理、共享、数据调度等方面充分体现智能化和自动化。

➤ 可靠性

浪潮舆情云监测系统具有很高的可靠性。采用稳定、成熟的先进技术，以及高质量的设备。相关工作人员在系统运行过程中以及进行系统安装时，指出可能出现的系统故障及具体排除的方法和手段，确保浪潮舆情云监测系统 7×24 小时无故障工作。

➤ 可扩展性原则

由于随着新技术的发展，虽然现在的产品方案解决了目前问题，但是随着时间的变化，新类型的舆情信息的出现，原有的问题解决方案可能满足不了其需求，这时就需要对原有的产品进行升级。浪潮舆情云监测系统具备良好的可扩展性，能够随着新技术的发展以及时间的变化，及时进行产品的升级与更新。

➤ 先进性原则

采用当今国内、国际上最先进和成熟的计算机软硬件平台、软件设计编程方法、开放式的体系结构和信息安全保障体系，浪潮舆情云监测系统能够最大限度

地适应今后的业务发展变化需要。

➤ 安全性

不论是基础平台、还是各个业务子系统，整个浪潮舆情云监测系统的设计遵循了安全性的重要原则。在网络系统构建硬件方面，浪潮舆情云监测系统采用并发技术及相应的虚拟化存储等相应技术，从而保证了网络带宽可以满足该网络的实际需要。浪潮舆情云监测系统硬件选型以安全性、稳定性为主。浪潮舆情云监测系统具有检错、纠错能力，具有完善的备份措施。在系统出现故障时，能够在较短的时间内恢复系统运行。浪潮舆情云监测系统能对网络和设备的运行情况进行监测和记录。

➤ 实用性和最优性价比原则

在整个系统设计中，浪潮舆情云监测系统从实际出发，切实满足用户需要，保证实现本项目和本工程的建设目标。同时，本产品作好了适度问题解决与经济投入的比较，注重实用，有较高的性能价格比。

2.3 浪潮舆情云监测系统架构

2.3.1 信息采集

在浪潮舆情云监测系统中，用户可自定义监测站点及监测关键词，采集范围包括：新闻、论坛、微博、贴吧、博客、视频、报刊杂志等。结合元搜索技术（搜索引擎结果补充），确保信息全面性。

定向抓取源：浪潮舆情云监测系统对相关站点中的新闻、论坛、微博、贴吧、博客、视频、报刊杂志等网络媒体实现全面的抓取，同时支持对主流新闻网站分页、评论内容的采集以及对论坛点击数、回帖数的抓取。

搜索引擎结果：独有的智能爬虫系统还可以自动跟踪多个搜索引擎的搜索结果，对监测数据进行补充。

自定义 URL 来源及采集频率：除系统常规监测范围外，用户可以通过设定采集的栏目、URL、更新时间、扫描间隔等，自定义需要重点监测的信息，以便及时发现目标信息源的最新变化，并以最快的速度将个性化关注内容采集到本地。用户可自定义监测站点及监测关键词，采集范围包括：新闻、论坛、微博、贴吧、博客、视频、报刊杂志等。结合元搜索技术（搜索引擎结果补充），确保信息全面性。

2.3.1.1 元搜索

浪潮舆情云监测系统采用用户设置的站点与关键词进行抓取，可以满足绝大部分的需求，但是总有部分信息在非重点的来源冒出来，对于这部分信息需要采用元搜索技术，即采用百度、Google 等作为采集资源，利用预先定义的语料搜索词库从同一个检索入口对多个资源提交检索请求。

浪潮舆情云监测系统是采用网络爬虫技术和网络信息挖掘技术，对网络各类信息进行汇集、分类、整合、筛选，以形成对网络热点、网民意见等的实时舆情报表，为决策层全面掌握舆情动态，做出正确舆论引导，提供分析依据。

搜索引擎范围广泛但是深度不够，比如新闻评论、大部分论坛的回复帖子、微博目前搜索引擎还不能抓取，结合浪潮舆情云监测系统互联网信息采集与搜索引擎的产出是一个完美的解决方案。

由于搜索引擎采集不能像自定义站点那样预先定制页面结构信息，对于这部分信息智能采用智能分析页面结构，进一步提取页面的元数据。信息采集后可与

采用一般方式获取的信息一样展现给最终用户。

功能特点：

➤ 高度的可扩展性

浪潮舆情云监测系统体系结构灵活支持，可灵活扩展新的抓取方式，比如，假设新出现推特微博，现有方式不能采集信息，可通过配置，结合系统提供的 js 引擎可产生新的信息采集方式，在第一时间适应新出现的站点。

➤ 增量检索

即使对于元搜索系统也支持增量式检索，所有信息不会重复检索。

➤ 智能页面结构分析

智能分析搜索引擎返回信息种类，智能提取页面信息。

➤ 广泛的元搜索源

目前浪潮舆情云监测系统支持百度新闻、百度贴吧、爱问知识人、天涯、百度论坛搜索、百度博客搜索、雅虎新闻、百度新闻搜索、搜狗新闻搜索、搜狗博客搜索、SOSO 新闻、SOSO 博客、有道新闻 RSS、有道新闻、有道论坛、Live 资讯、即可新闻 RSS、奇虎论坛搜索、奇虎博客搜索、新浪新闻搜索、即刻新闻搜索、新浪博客搜索、搜搜新闻、盘古新闻搜索、360 新闻搜索等。

2.3.1.2 贴吧、论坛采集

浪潮舆情云监测系统全面支持论坛、贴吧信息采集，浪潮舆情云监测系统支持论坛各板块的信息采集，目前支持一次性采集。一次性采集与新闻类似，每个帖子采集后不在跟踪，下次采集只访问新增的主贴。这种策略的好处是性能高，主要内容一般不会漏掉，因为重要内容一般是在主贴发布。

支持 SS0，国内大的论坛部分会跨多台服务器，系统支持在多个论坛间实现 SS0 访问，虽然此功能对用户来说不明显，但是在访问论坛时可以避免了重复登陆，优化了性能。

浪潮舆情云监测系统支持封闭论坛采集，即支持采集需要注册账号论坛中的信息。

浪潮舆情云监测系统整体配置简单快捷，方便用户维护，支持可视化配置。

2.3.1.3 博客、微博采集

浪潮舆情云监测系统可对博客进行抓取，也可和论坛一样持续跟踪。

浪潮舆情云监测系统同样支持一次性采集，一次性采集即仅抓取采集时刻的网页快照。

使用独特的 js 脚本引擎解析。支持新浪、腾讯、搜狐、网易、人民、凤凰、移动、天涯、和讯、Twitter 等微博平台的全网抓取，数据时效性在 5 分钟之内。

浪潮舆情云监测系统支持微博图片采集，如果同时发布了图片可以与文字一起采集。

2.3.1.4 数据存储

作为对采集信息的信息加工服务之一，浪潮舆情云监测系统支持采集网页之后提取其正文内容和智能摘要，在此基础上可以选定一个或多个信息批量导出到 Word、Excel、文本文件等文件中，从而更好的被信息处理人员利用。

浪潮舆情云监测系统支持保存采集的信息的快照，相关图片，以及全文索引。浪潮舆情云监测系统支持在多台服务器保存数据，数据存取采用分布式存取方

式，通过对分布式存储的支持，使系统具有了更大的横向扩展能力，从而支持保存更长时间的数据。

2.3.1.5 功能特点总结

维护人员可方便指定采集的目标站点或页面。

采用多线程并发搜索技术，采集过程高效准确。

提供高效的更新手段，已经采集过的信息不会重复采集，更新时只获取前次采集后更新的网页。

维护人员可方便设定信息监测的时间周期。

采集源定义，分解为网站名和频道名，使来源属性更为准确。

支持对采集的网页信息进行自动分类。

自动过滤网页中新闻的正文内容，剔除垃圾信息。

自动提取网页中的属性信息，包括时间、来源等。

支持对网页中表格内容的自动判断和获取。

对于论坛内容，能够对主帖和回帖分别抓取，并可按需要进行合并和关联，论坛内容一般还包括标题、关键词、作者、来源、时间、评论数、回复数等关键信息。

在海外部署有服务器，可对海外重点网站实施 7*24 小时全面监控

可以采集网页中的图片，并在正文中标记图片的位置；

采集过程成熟可靠，容错性强，完成初始设定后可长时间稳定运行。

提供数据导出和数据导入功能，可实现与其它系统文件级的信息共享，并为实现与其他系统数据交换提供了技术支持。

增加了线程的采集延时设置，实现对采集目标的采集强度控制。

采集程序内存管理更为科学，加强了运行的稳定性和可靠性。

2.3.2 信息处理

垃圾信息过滤：基于机器学习的垃圾过滤机制可以自动过滤广告、水贴等无效垃圾信息。

智能去重：采用“文章相似性技术”，根据文档内容的匹配程度确定是否重复、去重的级别；根据不同的需要特点分为：URL 去重、标题去重、正文去重三个级别。

HTML 内容提取：采用自主知识产权的 HTML 网页文本萃取技术自动提取任意复杂网页中的标题、内容、作者、发布时间等信息，自动跟踪文章分页；对于论坛信息自动分析主贴、回帖以及作者等信息。

快照保存：对于每个经过抓取和处理的网页，系统都存有一个纯文本的备份，方便用户快速浏览，也方便用户查看被删除的文章或帖子。

2.3.2.1 内容去噪

浪潮舆情云监测系统支持智能去噪与精确去噪。智能去噪可以大大减少配置工作量，如对搜索引擎返回的页面的分析。智能去噪支持基于页面分析的与基于多页面统计的，可适应大多数网站，对于极少数不规范网站也可采用智能分析。精确匹配的去噪可以取得最好的效果，实现完美去噪，但是配置工作量稍大。

采用自动过滤技术，利用网页结构分析，自动分析有用的网页，自动提取元

数据，自动过滤掉不需要采集的网页和媒体文件，有效避免垃圾信息的下载以及对带宽的浪费。

利用自动过滤技术，滤除网页中广告、版权、栏目等无用信息，并能自动识别网页真正的标题（很多网页都以 New Page 来标识），同时保存网页中与正文相关的表格和图片，识别正文中嵌入的广告。



去噪后：

雪碧麻辣周三携手招商银行，共推周三半价优惠

前不久，清华大学行为与大数据实验室发布了“清华幸福中国指数”，报告发现，中国人在周六的正面情绪最高、负面情绪最低，而以周三的正面情绪为最低、负面情绪最高。在社交网站上，“周三综合症”也经常成为搜索热词。不过面临“周三综合症”弃疗边缘的上班族们大可放心，在这个吃货经济时代，食物的力量不容小觑。各地美味在舌尖绽放，满足了胃更满足了心。而要说最能让人感到痛快减压的美食，麻辣二字便不得不提，我们熟悉的辣椒、胡椒、花椒，各个提神醒脑，吃完一身舒畅还唇齿留香。

为了提高大家的生活幸福指数，深受年轻人喜爱的“雪碧”品牌也于近日推出“麻辣周三”活动，邀请大家在周三这个前不着休息后不着周末的日子里，约上三五好友齐聚招商银行指定合作麻辣美食餐厅，不仅能享受半价优惠更能得到免费的1.25升装的“雪碧”饮料一瓶，让疲劳、压力随麻辣十足的料理尽情地释放，带来无限正能量！

麻辣解压！“雪碧”解辣！放火过来！

由于一般上班族平均每日有三分之一的时间处于工作中，工作状态的起伏变化将直接关系到人们的心理和情绪。所以一到周三，上班族会发现在紧绷而兴奋的神经持续2天后，工作量和信息量仍在不断增加，而下一个周末太远，情绪上就容易沮丧低落，产生“周三综合症”，进而影响对待工作的态度。甚至，心理上的焦虑不安也会影响生理上的肌肉酸痛、浑身乏力、消化不良等症状。所以不少上班族在意识到“周三综合症”后纷纷寻找有挑战有活力的活动，以便帮助自己的心理能量恢复到满格状态。

不过有趣的是，无论何种应对良方，相比最终大家都会同意，美食永远是无可取代的治愈法宝。今年，“雪碧”特别为需要解压的朋友们定制了“麻辣周三”，因为麻辣味的美食，不仅能够增进食欲，同时辣椒性温，能促进血液循环，帮助减轻肌肉和神经上的痛感。而拧开瓶盖、大口喝“雪碧”所带来的“透心凉，心飞扬”相信同样是不少人放飞疲倦情绪的不二法宝。如今，这股清新不仅能给被辣到不行的你灭火解辣，还能因与“重口味”的独特搭配，让减压后的心情更加痛快。

2.3.2.2 自动分类

浪潮舆情云监测系统采用自动分类技术，可以按统计进行分类，创建专有的分类模型，支持多分类。

浪潮舆情云监测系统支持基于统计的分类与基于规则的分类，基于统计或者说基于机器学习的分类，再准备好已分类好文章的基础上，系统对这些文章进行学习，统计出关键特征，在利用这些特征对文本进行分类，这种办法有点是自适应性，可适应未知情况，有新的情况，仅需修改语料，再次学习即可。

基于规则的分类，简单说是基于关键字的分类，可设置复杂条件，符合一定条件就归属某个类别。自动分类准确率达到 90%以上，满足大多数应用的实用要求。支持复分，一篇文章能分入多个符合条件的分类。

用户可以自由维护分类词表，人工添加或修改规则，词表大小没有限制。

用户可设置任意多个分类规则。

中国行业	2014年中国产业结构迎来跃升新契 [转载3条]	新闻	正面
新浪家居	京津冀三地广播携手推出环保主题系列 [转载2条]	论坛	正面
易车网	【苏州相城区二手君越】17.8万_	新闻	正面
新浪财经	小米牵手百度今晚独家首发移动电源 [转载3条]	论坛	正面
南昌新闻	北京春季房展开幕 环京楼盘参展数里 [转载44条]	新闻	正面
中商情报	2014-2018年中国休闲农业市	电子报	正面
人民网天	交管部门严查天津大道货车闯行 两小	新闻	正面

新浪论坛	偷天换日洋奶粉不姓“洋”，真品洋奶【转载2条】	论坛	负面
新浪论坛	偷天换日洋奶粉不姓“洋”，真品洋奶【转载2条】	论坛	负面
百度贴吧	【求助】电网网络玩网通区各种卡，求	贴吧群组	负面
新浪财经	【利安达】支撑金价上涨的因素还剩多【转载3条】	论坛	负面
新浪财经	【利安达】支撑金价上涨的因素还剩多【转载3条】	论坛	负面
新浪财经	【利安达】支撑金价上涨的因素还剩多【转载3条】	论坛	负面
新浪网博	商标侵权纠纷的诉讼代理、行政查处和	新闻	负面
新浪网博	2014年3月13日《	新闻	负面

正面信息与负面信息

2.3.2.3 自动排重

基于自然语言处理技术，浪潮舆情云监测系统根据文章内容相似程度计算相似文章，方便获取同一内容文章的所有传播网站。采用“文章相似性技术”，根据文档内容的匹配程度确定是否重复、去重的级别；根据不同的需要特点分为：URL 去重、标题去重、正文去重三个级别。在详细信息列表里，分为“过滤”与“不过滤”，展示相关文章。

2.3.2.4 自动摘要

浪潮舆情云监测系统支持文章的自动摘要技术，使读者可快速掌握文章内容。更重要的，通过文章的摘要，系统可更加准确的判断文章的中心内容。从而大大避免垃圾信息。

摘要：美国谷歌公司10日宣布，将于15日(本周二)将谷歌眼镜的购买计划“explorer program”面向所有美国居民开放。本次限时购买活动销售时间将从4月15日上午9点开始，谷歌眼镜零售价格依旧为1500美元，

新闻：谷歌眼镜周二发售 可穿戴热浪再起 的摘要

2.3.2.5 多语言支持

浪潮舆情云监测系统支持多种语言页面信息的采集，包括简体中文、繁体中文、英文、维语、藏语等语言。在数据库中网页信息应以统一的内码存储，因此对采集的网页信息需要统一进行内码转换处理。网页统一存储的内码可以选择 GBK，GB18030 或者 Unicode 等。不论选择哪种内码作为存储标准，都需要把各种网页内码统一转换成标准内码。

2.3.3 信息分析

文章权重计算：浪潮舆情云监测系统综合网站重要程度、文章出现位置、主题相关度、危机程度、点击回复次数、传播数量以及用户自定义规则等复杂参数计算的文章权重，加上基于自然语言处理技术的训练系统，能准确分析出重要舆情信息。

传播轨迹分析：浪潮舆情云监测系统可以对于重点舆情信息的舆情信息进行分析，展示重点舆情的传播时间、传播路径等。

舆情预警：浪潮舆情云监测系统可呈现预警信息走势和预警信息列表。建立多个舆情指标，对于突发舆情自动发出预警信号，在第一时间通过邮件方式通知用户，辅助进行舆情干预和引导。

媒体覆盖分析：浪潮舆情云监测系统可以对监测信息的媒体类型情况进行展示，同时对信息的主要传播媒体自动识别，进行整合分析，以图表呈现。

2.3.3.1 热点（主题）发现

利用相似性检索技术，通过转载、点击、评论、回复和报道率等数值，发现新闻热点、关键词热点、专题热点等。

热点发现的难点在于围绕一个主题的不同信息的聚合，即基于专题的热点分析排名，这种方式实现需要采用模型结合关键词的方式，模型方式不容易漏检，但建模复杂，关键词方式准确，但不容易覆盖全部内容，二者结合可以实现有益的互补。

2.3.3.2 传播路径分析

发现文章传播路径



2.3.3.3 网页快照

对抓取的页面, 系统生成快照保存在本地。



2.3.3.4 主题追踪

对于网络上的某一条信息, 可设定个人主题追踪线索, 指定过去和未来的追踪时段, 一方面提取以往信息, 另一方面当该信息有后续的相关报导的话, 会显示在定制的追踪栏目中。

正在追踪			更多
▶ 【正义网_】	三部门联合行动严...	03-28 18:22	🗑️
▶ 【新浪_股】	环保部在廊坊查出...	02-28 11:20	🗑️
▶ 【新华网_】	天津市临港经济区...	02-10 07:20	🗑️
▶ 【搜狐新闻】	圈子 - 天津临港...	02-09 22:40	🗑️
▶ 【中国经济】	天津女子买减肥药...	01-15 15:38	🗑️
▶ 【惠州论坛】	天津吉林安居房曝...	01-14 15:29	🗑️

2.3.3.5 监听预警

➤ 预警条件

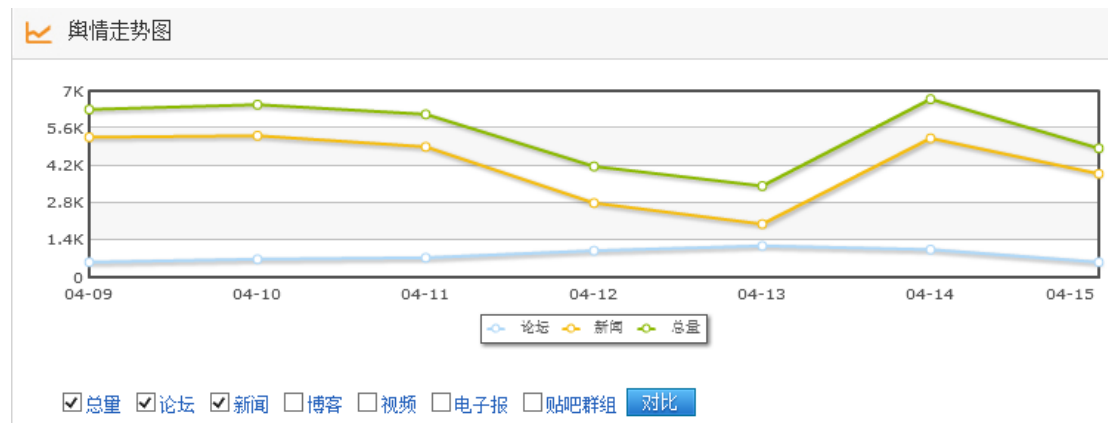
实时对网络上的舆情信息进行监测，新产生的舆情信息，或者已有舆情产生了新的转载，新的跟帖等变化，都会在第一时间采集过来，反映在系统当中，并及时通知舆情工作人员。自动发现热点，重点舆情通过站内消息、手机短信、邮件通知等方式预警。当出现突发的舆情信息时，需要在第一时间里通知不在现场的人员，在热点信息突发时，系统会自动启动短信发送任务，将该热点信息发送到指定用户/用户组的手机上。

➤ 预警方式

系统内建一系列敏感词表，当分析后的舆情信息的敏感词权重超过一定阈值后，系统自动触发舆情预警机制。预警包括对突发事件预警、涉及安全的敏感话题以及用户业务领域内的一些自定义预警。舆情预警方式有桌面（c/s）预警（对管理员弹出提示框体）、短信预警、邮件预警等提醒舆情分析人员。

2.3.3.6 趋势分析

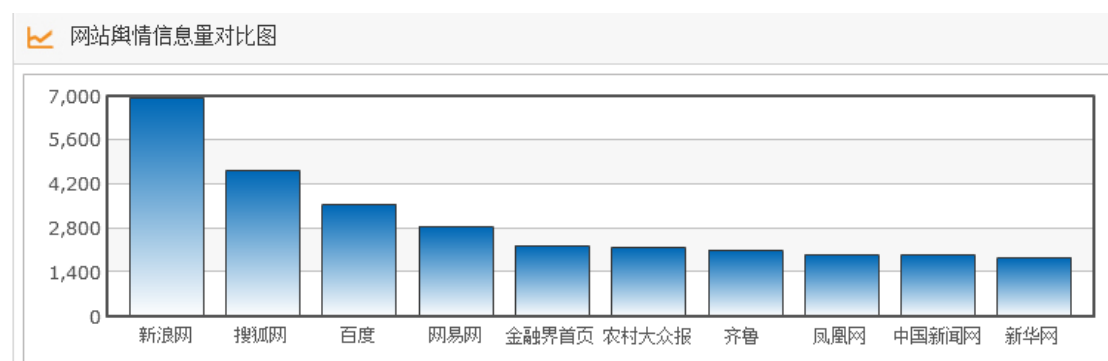
通过图表展示指定分类信息和时间的分布关系以及趋势分析，包括地域信息分布。以提供阶段性的分析。系统以图形化方式展示一段时间，在媒体上总的信息数量，也可支持某类信息在具体（如论坛）种类媒体上的发展趋势。



2.3.3.7 统计分析

➤ 站点分布

察看涉及到某种信息在指定站点上的信息情况, 包括总数, 分栏目数量等。使用户可在宏观上某站点的信息数量有总的了解。



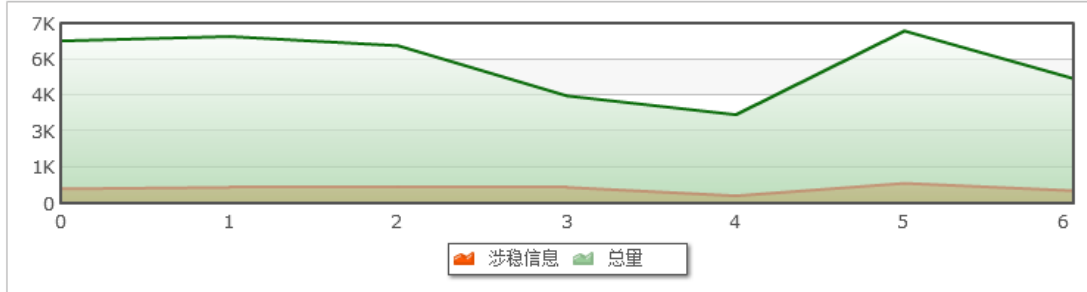
➤ 分类信息

分类统计, 提供各分类的信息数量, 使用户把握各分类的信息的活跃程度。

时间：今日 | 昨日 | 最近7天 | 最近30天 | 自定义

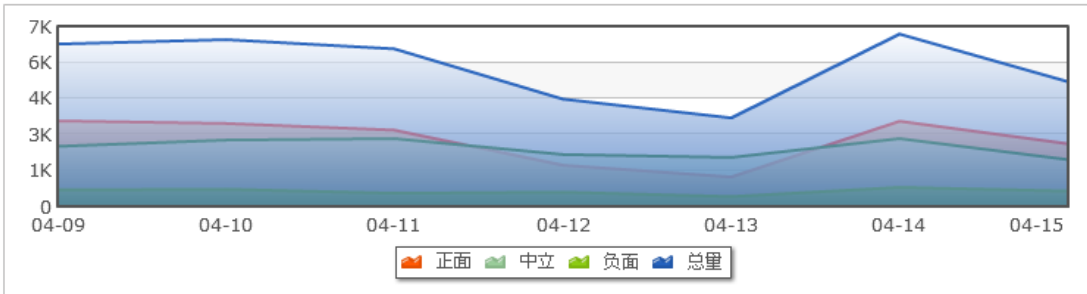
对比分析项目：总量 | 涉稳信息(全部) + 添加

舆情类别对比分析图



➤ 站点正负面信息统计

正负舆情走势图



2.3.4 信息展示

WEB 客户界面：基于云计算模式，用户可以使用 WEB 浏览器随时登录浪潮舆情云监测系统，在客户界面对舆情状况进行全面的了解。客户界面包含：最新预警、舆情走势、最新微博信息、分类舆情、载体覆盖情况等主要内容，并以列表及图表等可视化方式呈现，方便客户查看。导航栏清晰明了，方便用户查看各种分类；支持关键词组分类、媒体类型分类、信息属性分类查看。

舆情分析报告：根据舆情分析引擎处理后的结果库生成报告，客户可通过浏览器浏览，并且可以导出成为 Word、PDF 等格式的本地文档。方便客户对不同时段的重点和热点以及对事件舆情数据的把握。

2.3.4.1 首页

浪潮舆情云监测系统首页，使用户可全面了解系统最新信息，同时给出近期趋势图与相应热点。

浪潮舆情云监测系统支持多浏览器，满足用户浏览器使用多元化趋势。

2.3.4.2 舆情报告

浪潮舆情云监测系统根据舆情分析引擎处理后的结果库生成报告，用户可通过浏览器浏览，提供信息检索功能，根据指定条件对热点话题、倾向性进行查询，并浏览 信息的具体内容，提供决策支持。

2.3.4.3 简报形式

浪潮舆情云监测系统按需要提供及时报，及日、周、月等不同周期的报告，不同报告说明如下：

及时报：当有重大突发事件时及时对事件主题进行报导，展示发展趋势、分布、传播链。

日报：当日热点信息描述。

周报：本周热点信息描述及统计图表。

月报：本月热点信息描述及统计图表。

标题链接:	国办发文明要求严厉打击利用网络侵权假冒违法犯罪-天津网		
来源:	天津网_滚动	发布时间:	2014-04-14
媒体类型:	新闻	信息属性:	负面
总转载量:	0	当日转载量:	0
关键词:	三中全会,犯罪信息,涉稳信息		
摘要:	国办发文明要求严厉打击利用网络侵权假冒违法犯罪 来源: 中国新闻网关键字: 秋冬种;专利代理条例;网络出版物;网络接入;国办发 作者: 2014-04-14 16:38 据中国政府网消息, 国务院办公厅日前印发关于2014年全国打击侵犯知识产权和制售假冒伪劣商品工作要点的通知。		
修改此条 删除此条			

2.3.4.4 简报模版定制

简报的及时报、周报和月报中包含大量图形信息,这些信息都将以模块的方式,在简报编辑过程中进行调用,可设定相关参数,如时间段、时间颗粒度,统计坐标,然后自动生成图形并插入简报文档中。

2.3.4.5 简报格式转换

为了方便打印输出,报告完成后,可以导出至 Word、excel 和 PDF 等文档形式。

2.3.4.6 简报推送

结合邮件系统和短信猫,可以实现以 E-mail 和短信等多种方式发送报告和突发性的及时报。

2.3.5 其他

2.3.5.1 舆情个性化展示门户

浪潮舆情云监测系统提供面向个人的可定制的个性化展示页面,可以将系统

发布的热点信息、专题信息、聚类、关联词语、预警、统计等各种信息进行组织和调整，可进行模块的增加、删除。

浪潮舆情云监测系统不仅支持 PC 端查看舆情信息，同时支持 Android 系统、IOS 系统，手持移动终端可随时查看舆情信息。

2.3.5.2 外部接口与数据导入导出

浪潮舆情云监测系统开放外部接口，实现信息共享。

浪潮舆情云监测系统支持分析结果和简报等数据的导出功能，并支持 word、pdf 等多种格式。

第三章 技术优势

3.1 技术成熟性

浪潮舆情云监测系统，采用了成熟的先进技术，对浪潮舆情云监测系统的关键技术在前期的工作中进行了大量实验和攻关及原型建立，并在客户环境中得到了较好的应用。而且，浪潮舆情云监测系统所选取的硬件平台和软件平台，是具有良好的技术支持和发展前途的成熟产品。

先进的云计算技术：网络舆情监测运营云平台与舆情客户端配套使用。

自主研发智能爬虫：自主研发的智能网络爬虫（spider）系统，不留痕迹。

多种信息监控，封闭系统也能监控：新闻、论坛、博客、微博、贴吧、问答、QQ 群，甚至需要登录才能查看的封闭论坛都能监控。

监控范围广：全网爬虫技术和重点网站定向监测。

垃圾信息自动过滤：无关信息、垃圾贴、广告信息自动过滤技术，让信息精准无误。

时效性强：重点网站 5 分钟即可发现信息。客户即时提醒，5 分钟内通告用户。

独有的信息权重算法：根据网站重要性、文章重要性、主题相关度、信息危机度等多种指标综合计算的信息权重，重要信息一目了然。

智能相似文章算法：标题和内容做了一定修改的文章也能自动判断为相似，并生成传播轨迹。

自动分类与情感判断：将自然语言处理技术应用于系统，对信息精准分类，

并自动做情感分析。机器学习技术让系统越用越智能。

3.2 业务需求可靠性

符合我国现行法律、法规的规定和要求；并且，有机地融合了集中和分布两种模式的模块化、平台化的应用软件，不仅可以满足业务方面的各种需求，而且，还能够较好地解决业务需求可能面临的扩充及变动，具有较好的通用性与灵活的扩展及升级能力。