啄木鸟信息挖掘分析系统

南京大学 | 自然语言处理实验室

南京大学自然语言处理实验室

产品介绍

由于网上的信息量十分巨大，仅依靠人工的方法，难以应对网上海量信息的收集和处理。“工欲善其事，必先利其器”，政府或者企业需要舆情IT平台，实现自动化的网络舆情预警与分析，解决网络舆情管理过程中的舆情采集、分析、表达、干预等难题，从而梳理和客观呈现互联网上的热点舆情，使政府或者企业利用互联网舆情，汲取民间智慧，变网络舆情由被动防堵，化为主动梳理、引导。为推动构建和谐网络环境做出应有的贡献。

啄木鸟信息挖掘分析系统，是将自然语言处理技术应用在舆论情报服务的一次创新。系统使用智能爬虫、文本处理及大数据挖掘技术，为客户提供网络舆情7\*24小时监测、预警服务，根据预定的关键词在第一时间发现舆情信息。同时，系统配备手机短信、邮件实时提醒高危舆情信息，弥补了传统舆情监测在空间和时间上的不足。系统采用单点登录、手机绑定等方式的登录验证，数据库加密传输，多台从节点数据库服务器操作，保证系统安全性。

啄木鸟核心功能

舆情预警

舆情分析

系统通过对全媒体全面扫描和采集，情感分类自动判断正负面，发现舆情，系统（或人工）会推送"高危舆情"，并通过短信、邮件等形式进行舆情预警，重要舆情电话通知。

舆情监测

全面监测国内外网站、论坛、博客、微博、报刊、广播电视台等几百个采集点、关键词超500万个，1000多个信息库，重点采集与用户关注点有关的网络信息，可提供政府、行业、企事业等系统舆情监测产品；可实现舆情信息、事件、任务栏目分模块查阅。

舆情报告

舆情分析包括实时监测、媒体分析、传播路径、舆情报表等内容，分别从舆情时间分布、观点倾向、渠道传播情况三个角度对舆情进行数据化处理，并自动生成便于查阅的统计报表或图表，为舆情管控工作提供参考素材。

使用大数据挖掘及文本抽取技术，对信息进行数据分析和即时统计，系统可自动生成舆情报告产品体系。提供日/周/月/季/年分析报告、热点事件专题报告、分析预判舆情走势、提供应对建议。

SNAS系统特色和技术优势

**技术优势**

* 智能化的中文信息处理
* 分钟级舆情响应
* 采用分类、聚类等中文智能处理技术
* 智能数据分析引擎
* 可视化分析结果
* 自学习自增长监控目标
* 支持敏感词过滤提取
* 支持点击量、回复量采集
* 支持博客、论坛、贴吧及热门程度采集
* 支持网页编码自动识别
* 支持正文自动识别抽取
* 灵活的调度算法
* 可设置采集频率
* 支持分布式网络架构网站
* 全面的敏感信息分布式检索系统
* 合理的模块化结构及方便的分类监控
* 支持网页模板个性化配置
* 支持自动舆情图标生成
* 支持自动生成报告

全面性

全网监控：系统内置几百个国内主流网站采集源。

监控方式以定向监控与搜索引擎监控互补的方式；监控渠道包括（6种）：新闻、博客、论坛、微博、报纸、新闻网站这些保证了我们基础数据的全面性，也为后面的各种统计分析提供保障。

准确性

搜素引擎关键字采集、基于模板的定向采集、网页内容自动抽取、分析过滤广告等无效信息，以及高准确度文章情感倾向性判断等保证信息准确性。同时，利用智能信息检索和深度学习等技术保证信息获取的准确性和召回率。

及时性

超强搜索能力和并发能力，可设置网站检索更新周期，每分钟自动收集网络关键信息，先进的预警功能，舆情报送功能，使用户第一时间掌握关键舆情。