基于知识库和主题爬虫的南海舆情实时监测研究*

丁晟春 龚思兰 周文杰 王曰芬

(南京理工大学经济管理学院信息管理系 南京 210094)

摘 要 [目的/意义] 为满足对网络舆情的系统研究需求,需要将领域知识库作为重要支撑应用于相关研究中。[方法/过程] 首先,研究针对南海问题构建多语种南海舆情监测基本本体,基于该本体实现主题爬虫对舆情信息的采集;接着,基于优化的行块分布正文提取算法实现对舆情信息的正文抽取,获取关键字和摘要;最后,利用HTML5对舆情信息分析结果进行可视化展示。[结果/结论] 用户可根据具体需求利用构建的舆情监测系统对舆情信息实现系统的采集、处理和分析。

关键词 领域知识库 主题爬虫 网络舆情监测 南海问题

中图分类号 G353 文献标识码 A 文章编号 1002-1965(2016)05-0032-06

引用格式 丁晟春,龚思兰,周文杰,等.基于知识库和主题爬虫的南海舆情实时监测研究[J].情报杂志,2016,35(5):32-37.

DOI 10.3969/j.issn.1002-1965.2016.05.007

Research on Network Public Opinion Real-time Monitoring of the South China Sea Issue Based on Knowledge Base and Focused Crawler

Ding Shengchun Gong Silan Zhou Wenjie Wang Yuefen

(Department of Information Management, Nanjing University of Science and Technology, Nanjing 210094)

Abstract [Purpose/Significance] In order to meet the demand of systematic research in network public opinion research, we need to apply the corresponding knowledge base to it. [Method/Process] Addressing the South China Sea issue, this research built the multilingual knowledge base for public opinion information collection firstly, and then extracted the web page text by the optimized text extraction algorithm based on the distribution line block function, thus accessed to key words and abstract. Finally, the visual display of the result was implemented by using the API provided by HTML5. [Results/Conclusion] Users can use the public opinion monitoring system to achieve the public opinion information collection, processing and analysis.

Key words domain knowledge base focused crawler network public opinion monitoring the South China Sea Issue

0 引 言

中国与其它南海周边国家的关系中,南海问题是一个复杂而又重要的问题。南海问题涉及到了多个方面,如国家主权、历史、法律及当今敏感的国际环境,是国际关系学、国际法学、历史学、政治学等多个学科专家学者共同关注的焦点。网络舆情形成迅速,对社会影响巨大,使网络舆论成为社会舆论的一种重要表现形式。通过对与南海问题相关的网络舆情进行实时监测,可以及时把握南海舆情动向,对相关突发情况做出快速响应和处理,并为未来的南海问题对策提供借鉴参考。

收稿日期:2016-02-27 修回日期:2016-04-06

基金项目:国家社会科学基金项目"基于社会网络分析的网络舆情主题发现研究"(编号:15BTQ063);国家社会科学基金重点项目"大数据环境下社会舆情与决策支持方法体系研究"(编号:14AZD084);中央高校基本科研业务费专项资金资助(编号:30916011330)。

作者简介:丁晟春(ORCID:0000-0002-4269-021X),女,1971年生,硕士,副教授,研究方向:数据挖掘、知识工程;龚思兰(ORCID:0000-0002-1596-0560),女,1990年生,硕士研究生,研究方向:数据挖掘、知识工程;周文杰(ORCID:0000-0001-8599-3221),男,1991年生,研究方向:数据挖掘;王曰芬(ORCID:0000-0002-7143-7766),女,1965年生,博士,教授,研究方向:知识管理、竞争情报。

对南海问题进行舆情监测,需要整合领域知识库构建、网络信息采集、自然语言处理等众多技术,通过对互联网上大量网页信息的自动搜集、自动分类、自动聚类、统计分析,实现网络舆情监测、网络舆情预警等信息需求,形成图表、分析报告等结果,为作出正确的舆论引导提供分析依据。本文在对与网络舆情监测相关研究进行学习和总结的基础上,构建多语种南海问题知识库结合主题爬虫,实现南海问题相关网络舆情信息的采集、预处理、分析、监测和展示功能,系统架构明晰,操作简单,有助于科学高效地解决南海问题的网络舆情管理和预警。

1　相关工作

现有网络舆情监测研究主要集中于对网络舆情监测方法及其评价体系的构建研究等内容。

在网络舆情监测评价体系的构建研究中,张一文等通过建立指标体系来衡量和评价非常规突发事件网络舆情热度,为舆情管理提供理论依据[1]。王青等人根据现有的网络舆情监测与预警指标体系对网络舆情的传播特性、主题特征、内容价值等方面进行研究,提炼出网络舆情的大部分监测点[2];还对现有网络舆情监测指标体系进行整理与归纳,构建了更为科学系统的网络舆情监测与预警指标体系[3]。张玥等人从信息生命周期视角尝试拓展网络舆情监测的评价指标,构建了基于信息生命周期的网络舆情监测三维立体框架[4]。冯江平等从网络舆情主体相互作用的角度,构建网络舆情评价指标体系,并结合具体案例实现对所构建的指标体系的实践运用[5]。刘彬等人对舆情指标体系和分析模型的构建背景进行分析,提出了重大舆情分析指标体系,并构建了重大事件舆情预警分析模型[6]。

在网络舆情监测方法的研究中,陈忆金等以新闻评论、论坛帖子、BBS 等数据为研究数据来源,对网络舆情信息相关监测技术进行探索性研究,并将成果应用在网络舆情监测系统中[7]。唐涛研究了基于网络文本挖掘的监测模型和基于搜索日志挖掘的监测两种网络舆情监测方法,并分别进行了案例分析,取得了较好的效果[8]。Talvis 等人通过对 twitter 信息进行语言和统计分析实现对流感疫情的实时监测,获得了 92% 的危重病例自动识别准确率[9]。Velardi 等人开发了一种自动学习算法,采用症状驱动的方法利用医学语言模型,挖掘 twitter 信息,实现对流行病流行趋势的有效监测[10]。Lauschke 等提出了一种基于主题的用户分析和监测方法,用于舆情内容演化的检测和监控,并用 twitter 上的数据评估该方法的性能[11]。Robert 等提出了一种用户可配置的监控系统对 twitter 上有关火灾的推文进行实时监测[12]。Terpstra 等人通过应用程序对提取的信息进行运营危机管理,自动过滤无效信息,实现对 Twitter 的实时危机分析[13]。

目前有关网络舆情的研究还缺乏相应知识库的支撑,若要对网络舆情进行系统的分析,必须要有一个完整的、全面的、有代表性的知识库支撑[14]。陈越等在分析建立网络舆情知识库必要性的基础上,对各类知识库在数据采集、信息抽取、话题检测与追踪等网络舆情监测与预警过程中的作用以及知识库间的关联关系进行了研究[15]。

主题网络爬虫是一种利用在采集页面过程获得的特定信息下载特定主题网页的程序[16]。网络舆情监测一般是面向行业监测,倾向于使用面向主题爬虫,根据不同的采集需要,添加不同的中间件用于过滤无效的 URLs 及无效或不需要的网页信息,只存储所需的部分[17]。李月超等人分析了主题爬虫技术特点及处理流程,指出在现有网络舆情监测背景下,应加强网络舆情监控系统主题网络爬虫的功能研究以满足面向特定范围内的信息采集和监测要求[18]。杨贞通过将本体知识库与主题爬虫相结合,提出了一种基于本体的主题爬虫最好优先爬行算法,设计并实现了主题爬虫原型系统,实验结果表明该系统有效扩大了主题爬虫的搜索范围[19]。

基于上述研究,本文提出了一种基于知识库和主题爬虫的网络舆情监测方法,通过将领域本体知识库与主题爬虫相结合,用以扩大主题爬虫的搜索范围并提高其搜索精确度。在此基础上建立的网络舆情监测系统可以获得相对完备和精准的知识体系及数据支撑,最终实现对网络舆情的高效预警和管理。

2　南海问题领域多语种知识库的构建

2.1　知识库构建思路　南海问题主要涉及南海沿海国中国、菲律宾、文莱、马来西亚、越南和印度尼西亚,近来由于美国等大国的力量介入,使得南海问题涉及的主体开始多元化,产生的舆情信息资源也呈现出多语种化。本文搜集与南海问题相关的多语种舆情信息资源及文献信息资源,根据南海问题涉及的国家信息、地理信息不同,收集包括汉语(中文)、英语、越南语、菲律宾语、马来语、印度尼西亚语等在内的多语种信息资源,并进行整理、归类,总结了目前存在的主要舆情信息资源;在确定南海问题性质的基础之上选择合适的上层本体结构——事件本体,确定南海问题的设计标准;在知识库的建模过程中,采用本体的思想进行南海问题知识结构的挖掘。

知识库的逻辑结构分为本体库和数据库两大部分,形成南海争端事件本体、国家信息库、地理信息库、

关注方信息库、状态信息库等本体库，以及舆情信息库、文献信息库、观点库等数据库协同运作的知识库结构。知识库构建思路如图1所示。

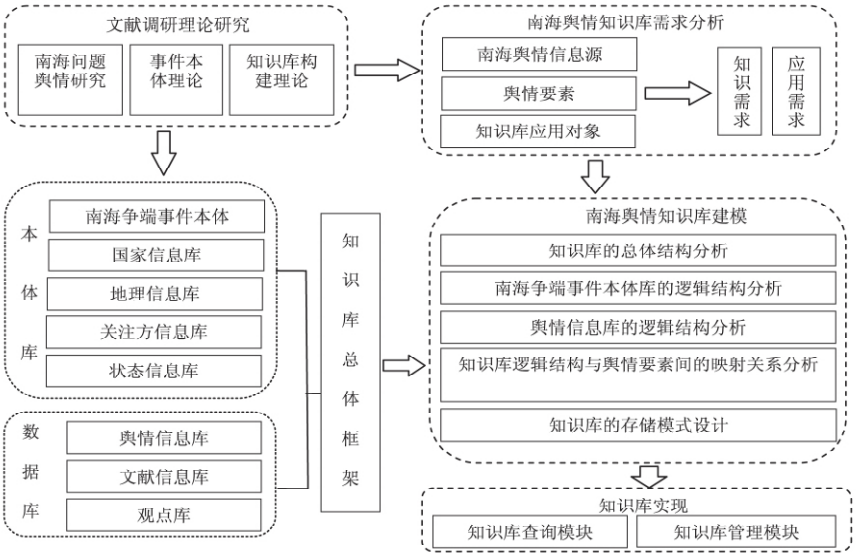

图1　南海问题领域多语种知识库构建思路

2.2　南海舆情本体库的构建　本文所构建的本体库以南海争端事件本体库为核心，包含南海争端类（过程类、事件类），国家类（利益主体类、一般国家类），关注者类（个人类、研究机构类、媒体类、政府类）和岛屿类四个基本大类。同时还建立了起因、包括、是……的过程、继……之后、关注了、被关注等15种类之间的相互关系，能够较为完整地描述南海争端事件，符合南海问题知识存储的实际需求。图2为类之间的部分类关系的展示。

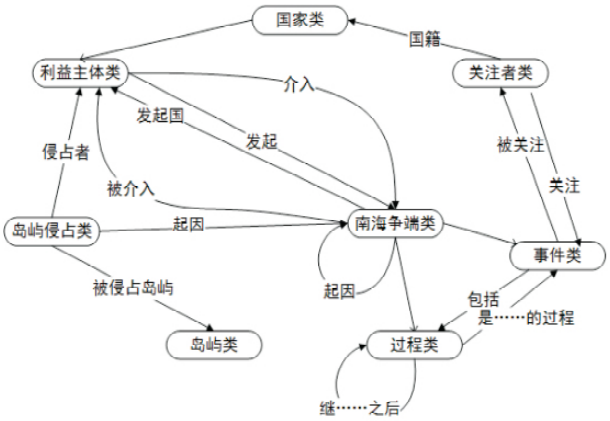

图2　南海问题本体库的类关系

例如，事件类和过程类之间的［包括］和［是……的过程］是一组互逆关系，表示事件和过程之间的关系，即事件由多个过程组成，所以是事件包括过程，而多个过程的发生形成了整个事件，所以说过程是事件的过程。类关系［起因］表示南海争端之间的因果关系，表现为南海争端类与南海争端类自身的关联，能够为事件类和过程类继承。本文通过protégé进行了相关的类、属性和关系的建立，实例的添加是通过所构建的知识库管理系统来完成的，图3为实例添加页面。

图3　关注者类实例录入界面

3　基于主题爬虫的网络舆情信息实时采集

3.1　基本思想的设计　舆情信息的采集主要通过主题爬虫实现，在所构建的南海问题领域多语种知识库的支持下，根据系统用户提供的特定关键词和网站地址进行相关主题内容爬取。系统会根据用户提供的关键词，与系统后台链接的南海领域知识库进行匹配，对与该关键词对应的多语种知识内容统一进行抽取，获得相关主题词群，系统根据该主题词群实现对该关键字涉及的所有相关主题内容的全面爬取。例如，系统用户提供的关键词为岛屿名称“中沙群岛”，系统后台会根据该关键词与后台知识库进行关联匹配，获取与“中沙群岛”对应的多语种关键词“the Zhongsha Islands”“Kepulauan pasir”“Quânđảo cát”“Kepulauan Sand”等，再进一步根据上述关键词抽取与该岛屿名对应的涉及争端国家信息库，文献信息库，下行地理信息库（中沙大环礁［西门暗沙、本固暗沙、美滨暗沙等］，“北部大陆架［西门暗沙、本固暗沙、美滨暗沙等］”，“中央海山［黄岩岛［民主礁］、宪法暗沙、中南

暗沙、中南海山、龙南海山、长龙海山等]") 中所涉及的主题词群。主题爬虫根据获取的主题词群,抓取符合条件的舆情信息,该方法能够有效扩大主题爬虫的搜索范围并提高其搜索精确度。

在信息采集过程中,需要进行无效信息的过滤、内容页的识别、内容去重。系统从所抓取的符合条件的页面中提取所需数据项,提取的数据项主要为标题、正文、发布日期等。同时,还需根据网络动态改变网页爬取速度,并不断切换模拟的浏览器的 user agent 信息,清洗 HTML 并采用正文抽取方法获取正文内容,并根据关键字匹配决定是否持久化到数据库中,为后续的舆情监测和分析提供数据基础。

3.2　关键技术　主题爬虫方面使用 Python 的 scrapy 框架定制 Spider 进行爬取,采用单独调用脚本或从前端触发爬虫的运行。在信息采集过程中,主题爬虫中对所有的 drop item(指定项) 进行了异常配置,抓取时可实时显示 item(对应项) 的处理结果。每次爬虫起止均有日志信息记录,记录在根目录下的 scs_crawler. log 文件中,但未对日志文件大小进行控制,对基于 scrapy 的爬虫,配置 graphite 来实时监控爬虫的采集情况。整个过程主要涉及 URL 去重、内容页识别、网页的预处理、基本数据项提取、正文提取等。

3.2.1　URL 去重　爬虫爬取网站的过程中,需要对已爬取过的链接进行识别,这将直接影响到爬虫的爬行效率。系统选择采用基于内存的方式爬取网页信息,爬虫从数据库中读取已下载的网页 URL,导入到 HashTable 中进行 URL 列表的初始化,后续下载过的每个网页(包括最终未存入数据库的网页) URL 均存入 HashTable 中,用于下一个 URL 的过滤。

3.2.2　内容页识别　系统选择根据 URL 的规则和根据正则表达式两种方式进行内容页的识别。根据 URL 的规则识别主要基于域名分级、是否包含日期,辅助以页面中是否包含日期、页面是否过期。在配置起始网站时,可进行内容页地址的正则表达式配置,且支持多个,中间用“|”作为分隔符存入数据库中。网站配置内容页的正则匹配后,爬虫主要根据正则进行内容页的识别。

3.2.3　网页的预处理　网页预处理包括编码问题的处理、链接抽取和过滤、无效信息的去除。

a. 编码问题处理: 识别网站编码,根据 HTTP Response 的 content encoding 提取编码,根据编码解码为 Unicode 进行后续所有处理,符合抓取条件的存入数据库,输出时在 HTTP Response 中设置编码换为 UTF－8 编码,即可正确显示在 charset 为 UTF－8 的 HTML 中。

b. 链接抽取和过滤: 链接抽取需过滤无效链接,设定过滤规则,包括对相对路径和绝对路径进行处理,加入域名地址或过滤不同域名下的链接。

c. 无效信息去除: 为提高数据项提取的精度,需去除 HTML 中的无效信息,如脚本、样式、链接的 URL 等。系统使用 python 的 BeatifulSoup 库进行 html 处理,具体措施包括去除评论、删除链接、去除内嵌的 iframe 框等。同时,还通过丰富 HTML 标签、将 DOM 树和正则匹配相结合,使得系统对非中文的无关噪声进行的处理更为全面。

3.2.4　基本数据项提取　网页预处理后需要进行数据的持久化存储,需要字段为网页地址、采集日期、标题、正文、发布时间,其中需要抽取的基本数据项为标题和发布时间。

标题的提取主要根据网页的 title 属性,以及网页内部的 h1 和 h2 标签中的文字,将 h1 及 h2 分别与 title 进行匹配,如果匹配成功则直接返回 title; 发布时间的提取根据正则进行匹配,分为日期和时间两个部分进行匹配,最终的时间戳由两个部分进行合并,不存在发布日期则 drop。

3.2.5　正文提取　经过了以上几步的过滤,此时的页面默认成为可提取正文的内容页。系统采用的方法来源于哈工大的基于行块分布的正文提取算法[20],是一种基于文本密度的方法,处理步骤如下: a. 清洗 HTML,由于链接在预处理阶段已剔除,对正文提取的影响将会减少; b. 计算每行的数值,并计算出最大正子串的起止位置; c. 根据 2 中得出的起止位置返回正文行块,并去除 HTML 标签和空行; d. 系统利用 Python 实现抽取计算起始和结束行号,并返回正文行块。

4　系统设计与应用

系统前端采用 JS 的 MVC 框架 Angular JS,整合 HTML5、CSS3、D3. js 进行数据交互和显示,使用 AJAX 与后端进行通信。后端采用 Python 的 flask web 框架提供符合 restful 规范的 API 来响应 AJAX 请求,返回相应的数据,并使用多个 Python 模块进行一系列的数据处理。

4.1　系统功能设计　系统主要由信息采集模块、信息分析处理模块、数据可视化展示模块构成,具体功能结构如图 4 所示。

4.2　舆情信息分析结果展示　舆情信息分析处理包括关键字和摘要抽取、地理位置分类两部分。系统主要基于词频统计(TF－IDF) 来进行关键字和摘要的抽取,根据建立的南海地理位置类目,进行舆情信息地理分类,并给信息加上分类标识。系统的地理位置分类,关键词和摘要抽取等功能均提供接口供用户根据需求修改。信息的分析处理在后续的页面操作中进行

触发，系统后台提供API供前台获取数据，前台将数据使用不同的形式(表格文字、图形等)展示出来。

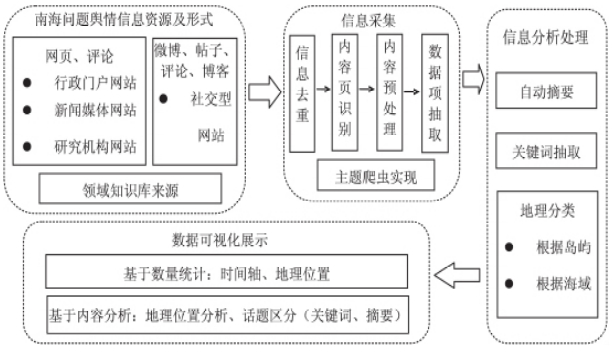

图4 系统功能结构图

数据的维度有来源、数量、时间等，系统选择基于内容和基于数量两种方式实现舆情信息分析结果的可视化展示。其中，基于内容的文章列表信息表格数据展示形式如图5所示。

图5 文章列表表格数据展示(越南语)

基于数量的结果数据可视化展示方式包括区域图、折线图、柱形图、堆叠面积图等多种方式。系统中数据展示都以基于后台返回的JSON数据为基础，图表数据生成使用基于D3.js的nv.d3.js进行展示，本文将nv.d3.js中的堆叠面积图封装成Angular JS的指令，如下:

```
appDirectives.directive('areaChart', function() {
function link(scope, ele, attr) {
var ele = ele[0]
var svg = d3.select(ele).append('svg')
var colors = d3.scale.category20();
var keyColor = ;
var chart = nv.models.stackedAreaChart()
.useInteractiveGuideline(true)
.x(function(d){return d[0]}) //以key作为x轴数据来源
.y(function(d){return d[1]}) //以values作为y轴数据来源
.color(function(d,i){return colors(d.key)}) //设置区域图的颜色
.transitionDuration(300);
chart.xAxis.tickFormat(function(d) { //格式化时间
return d3.time.format('%x')(new Date(d))
});
wrapChartData() //初次进行绘制
scope.$watch("data", wrapChartData); //监听数据变化
function wrapChartData() { //绘制图形
var data = scope.data //从作用域中的data重新获取数据
svg.datum(data).transition().duration(1000).call(chart)
nv.utils.windowResize(chart.update); //监听窗口的大小变化
}
}
return {
link: link,
restrict: 'E',//作为HTML元素使用，即 <areaChart> </areaChart>
scope: { data: '=' } //数据来源为元素中的data属性中的值
}
})
```

本文以堆叠面积图为例，对舆情信息演化状态进行直观展示，方便用户快速准确地获取舆情关键信息源、关键时间点等重要信息。横轴可选择发布时间、采集时间，也可进行时间跨度的选择(按天、按月、按年)；纵轴为文章数量，可选择按网站名称、网站类型进行分组统计。从图6中可以看出，2014年5月9日到5月10日，中华网发布的南海舆情信息大幅度上升，是由于“5·6菲律宾扣押中国渔民事件”的发生在随后几天内获得了国内媒体的广泛关注和持续报道。2014年5月18日到5月21日，越南网发布的南海舆情信息大幅度上升，是由于“5.13越南反华游行”致使中国政府于5月18号暂停部分与越南交往计划，引起了越南方面的注意，会引发对两国存在争端的南海问题的讨论。

图6 舆情信息源—采集时间－舆情数量分布图

5 结论与展望

本文构建了南海问题领域知识库，以期满足对南海问题领域基本信息的存储，实现数据共享，并进行一些自动的知识推理。在此基础之上，研究基于主题爬虫对满足条件的网络舆情信息进行实时采集，并对采集结果进行了初步的预处理和内容抽取。最终，基于上述研究成果构建了基本的南海舆情监测系统，并实现了信息采集、网页正文抽取、自动摘要抽取、数据可

视化等功能。

本系统现有不足之处主要包括:a. 构建知识库的成分较为简单,知识库内本体库的分类结构内容较少,而且整个知识库的构建缺乏足够的理论支撑,需要进一步的完善;b. 信息采集过程没有对 IP 进行动态变化,实现"礼貌"爬取方面有待加强;c. 爬虫监控需要完善,实时监控对象可为处理网站数量、保存的网页数量、内存使用率等,需要嵌入到页面模块中;d. 对提取正文、摘要时的准确度没有进行大量的统计分析,后续的准确度优化可以使用混合的方式进行正文、摘要等的提取,如使用页面模板等;e. 数据的展示可使用更加丰富的表现形式,维度可继续增加,在用户体验上有待优化。

网络舆情监测涉及到多个领域的技术,构建一个完整的网络舆情监测系统,除了在本系统中运用的知识外,还需涉及到分布式、聚类分析、情感分析等技术,如何进行有效的信息抓取和分析,让信息的处理更加智能是未来我们需要继续探索的地方。

参 考 文 献

[1] 张一文,齐佳音,方滨兴,李欲晓. 非常规突发事件网络舆情热度评价指标体系构建[J]. 情报杂志,2010,29(11):71 - 75,117.

[2] 王　青,成　颖,巢乃鹏. 网络舆情监测及预警指标体系研究综述[J]. 情报科学,2011,29(7):1104 - 1108.

[3] 王　青,成　颖,巢乃鹏. 网络舆情监测及预警指标体系构建研究[J]. 图书情报工作,2011,55(8):54 - 57.

[4] 张　玥,罗　萍,刘千里. 基于信息生命周期理论的网络舆情监测研究[J]. 情报科学,2013,31(11):22 - 25.

[5] 冯江平,张　月,赵舒贞,等. 网络舆情评价指标体系的构建与应用[J]. 云南师范大学学报:哲学社会科学版,2014(2):75 - 84.

[6] 刘　彬,董茜茜. 重大事件舆情监测指标体系与分析模型[J]. 统计与决策,2015(11):4 - 8.

[7] 陈忆金,曹树金,陈少驰,等. 网络舆情信息监测研究进展[J]. 图书情报知识,2011,144(6):41 - 49.

[8] 唐　涛. 基于情报学方法的网络舆情监测研究[J]. 情报科学,2014,32(1):124 - 127,137.

[9] Talvis,Karolos Chorianopoulos, Kostantinos Kermanidis,et al. Real - time Monitoring of Flu Epidemics Through Linguistic And Statistical Analysis of Twitter Messages[C]//Proceedings - 9th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2014(7):83 - 87.

[10] Velardi Paola,Stilo Giovanni,Tozzi Alberto E,et al. Twitter Mining for Fine - grained Syndromic Surveillance[J]. Artificial Intelligence in Medicine, 2014,61(3):153 - 163.

[11] Lauschke,Claudia, Ntoutsi,et al. Monitoring User Evolution in Twitter[C]//Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Asonam,2012:972 - 977.

[12] Power Robert,Robinson Bella,Colton John,et al. A Case Study for Monitoring Fires with Twitter[C]//ISCRAM 2015 Conference Proceedings - 12th International Conference on Information Systems for Crisis Response and Management, 2015.

[13] Terpstra, Teun, Stronkman R,et al. Towards a Realtime Twitter Analysis During Crises for Operational Crisis Management[C]//ISCRAM 2012 Conference Proceedings - 9th International Conference on Information Systems for Crisis Response and Management, 2012.

[14] 李　纲,陈璟浩. 突发公共事件网络舆情研究综述[J]. 图书情报知识,2014(2):111 - 119.

[15] 陈　越,李超零,黄惠新. 网络舆情监测与预警中的知识库研究[J]. 图书情报工作(增刊),2011(2):262 - 266.

[16] 于　娟,刘　强. 主题网络爬虫研究综述[J]. 计算机工程与科学,2015,37(2):231 - 237.

[17] 李　婧,刘志明,崔朝国,等. 基于微博的舆情监测与分析的研究[J]. 智能计算机与应用,2013,3(2):50 - 53.

[18] 李月超,李芸洁,李　勤,等. 网络舆情监控系统中主题网络爬虫的研究[J]. 电脑知识与技术,2015,11(2):46 - 47.

[19] 杨　贞. 基于本体的主题爬虫的设计与实现[D]. 合肥:合肥工业大学,2008.

[20] 陈　鑫. 基于行块分布函数的通用网页正文抽取法[EB/OL]. [2009]. http://code.google.com/p/cx - extractor/.

(责编:贺小利)

(上接第 51 页)

"福喜事件"为例[J]. 江西社会科学,2015(11):247 - 251.

[9] 桑　亮,许正林. 微博意见领袖的形成机制及其影响[J]. 当代传播,2011(3):12 - 14.

[10] Atefeh F,Khreich W. A Survey of Techniques for Event Detection in Twitter[J]. Computational Intelligence,2013,31(1):132 - 164.

[11] Bakshy E,Rosenn I,Marlow C,et al. The Role of Social Networks in Information Diffusion[C]//Proceedings of the 21st International Conference on World Wide Web,2012:519 - 528.

[12] 冯　锐,李亚娇. 社交网站中知识扩散机制及影响因素研究[J]. 远程教育杂志,2014(3):41 - 48.

[13] 相丽玲,王　晴. 信息公开背景下网络舆情危机演化特征及治理机制研究[J]. 情报科学,2014(4):26 - 30.

[14] 任立肖,张　亮,张春莉. 无标度网络机制下网络舆情传播演化规律分析[J]. 现代情报,2014(2):8 - 12.

[15] 周爱明. 南京"宝马肇事案"舆论传播的几点反思[J]. 传媒观察,2015(11):27 - 28.

[16] 任立肖,张　亮,杜子平,等. 复杂网络上的网络舆情演化模型研究述评[J]. 情报科学,2014(8):148 - 156.

[17] 张一文,齐佳音,方滨兴,等. 基于贝叶斯网络建模的习常规危机事件网络舆情预警研究[J]. 图书情报工作,2012,56(2):76 - 81.

(责编:刘影梅)