

Application of Image Search Engine in Public Opinion Guidance and Control

Kehan Lv, Wei Xu, Qinyu Wang, Xueyan Sun, Yiyu Zhang, Rongchun Wan*

Jiangsu Police Institute, Nanjing Jiangsu
Email: *2938163378@qq.com

Received: Apr. 1st, 2019; accepted: Apr. 12th, 2019; published: Apr. 24th, 2019

Abstract

In order to strengthen the construction of public security information, improve the ability of network supervision of public security organs, and solve the problem that public security organs can not accurately control the direction of public opinion in public opinion incidents, this paper chooses the function of image search engine to carry out experimental research. The analysis of the efficiency of image search engine by experiment, and the research on how to use image search engine to real-time monitor public opinion in specific public opinion events, provide experimental basis for better use of image search engine in practice and better public opinion guidance.

Keywords

Image Search Engine, Graph Search, Public Opinion Guidance and Control, Hash Value, Baidu, Sogou, 360

图片搜索引擎在舆情导控中运用的实验研究

吕柯汉, 徐伟, 王沁宇, 孙雪岩, 张一语, 万荣春*

江苏警官学院, 江苏 南京
Email: *2938163378@qq.com

收稿日期: 2019年4月1日; 录用日期: 2019年4月12日; 发布日期: 2019年4月24日

摘要

图片搜索引擎是新兴的一种专门用来查询图形、图像并以网站名称或内容为形式反馈检索结果的网络检索工具。通过实验对图片搜索引擎以图搜图检索效率的分析, 以及对具体舆论事件进行图片搜索引擎的

*通讯作者。

文章引用: 吕柯汉, 徐伟, 王沁宇, 孙雪岩, 张一语, 万荣春. 图片搜索引擎在舆情导控中运用的实验研究[J]. 图像与信号处理, 2019, 8(2): 72-82. DOI: 10.12677/jisp.2019.82011

检索研究,探讨了图片搜索引擎“以图搜图”功能在舆情导控中的运用方法。本文可以为图片搜索引擎“以图搜图”功能的运用提供新的角度和方向,为公安一线干警开展网络舆情导控工作提供参考。

关键词

图片搜索引擎,以图搜图,舆情导控,哈希值,百度,搜狗,360

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自进入互联网时代,信息资源被广泛利用,有力推动了社会建设和经济发展。随着人们对信息的搜集、存储、检索、分析、应用、评估需求越来越大,搜索引擎应运而生。搜索引擎为广大用户获得最准确、最广泛、最具时效性的信息提供了坚实基础。近年来随着以微信为代表的各类新媒体以锐不可当之势进入社会生活,图片、视频信息在互联网信息资源中上传播数量急剧增加,以关键词为主的传统搜索引擎已经无法满足人们的需求,图片搜索引擎的快速崛起已然不可阻挡。

图片搜索引擎是近十年来兴起的一种专门用来查询图形、图像并以网站名称或内容为形式反馈检索结果的网络检索工具。其中,最具代表性的功能便是“以图搜图”。“以图搜图”即相似图片搜索,又称基于内容的图像检索或反向图像搜索,是根据用户提供的目标图片,检索与其相似的图片等信息的一种专业搜索引擎。以图搜图技术不仅仅在商品比价、技术辅助、视频搜索等领域具有广阔的发展前景,它更可为公安机关进行网络监管、舆情导控提供重要技术支持。

现今,以图搜图技术的发展还不成熟,公安网络监管、舆情导控中鲜见运用,文献中也没有相关的研究。本文拟在实验基础上,通过对图片搜索引擎“以图搜图”的检索效率的研究分析,以及结合网络舆情典型事件搜索引擎“以图搜图”的检索结果的研究分析,探讨运用图片搜索引擎“以图搜图”功能实现舆情导控,为图片搜索引擎“以图搜图”功能的运用提供新的角度和方向,从而为公安一线干警开展网络舆情导控工作提供参考。

2. 实验设计

(一) 实验原理

图片搜索引擎“以图搜图”的工作原理就是通过特定的算法(如 SIFT 描述子、指纹算法函数、bundling features 算法)将目标图片进行特征提取,然后将所提取的图片特征信息进行编码,并将海量图片编码做查找表,通过相似度匹配运算,以目标图片的编码值为基础,在图片搜索引擎图片库中进行局部或者全局的相似度计算,根据所需图像的设定阈值,保留相似度高的图片,最后运用特征检测算法筛选出最佳匹配图片。

图片搜索引擎“以图搜图”的实现有一项关键的技术叫做“感知哈希算法”(Perceptual Hash Algorithm)。它的处理方法是每一张图片生成一个指纹(fingerprint)字符串,通过比较不同图片的指纹,计算它们的相似度,从而得出两张图片是否相似。具体有以下六个步骤:

第一步,读取所需图片。

第二步,缩小图片尺寸。将读取图片缩小到 8×8 的尺寸,即 64 个像素。目的在于只保留读取图片

明暗和结构等基本信息，同时去除图片细节，以及图片不同尺寸，比例带来的差异。

第三步，简化图片色彩。将缩小后的目标图片转化为 64 级灰度，即所有像素点只有 64 种颜色。

第四步，计算平均值。计算 64 个像素灰度的平均值。

第五步，比较像素的灰度。将计算出的平均值与每个像素的灰度进行比较，灰度大于等于平均值的，记为 1；灰度小于平均值的，记为 0。

第六步，计算哈希值。经上一步操作，将所记的所有数组合在一起，构成一个 64 位的整数，这就是读取图片的指纹字符串。根据得到的指纹字符串，对不同的图片进行对比。理论上，这与计算“汉明距离”等价。如果两张图片的指纹字符串不同的数据位不超过 5，则说明两张图片相似；反之，大于 10 则说明两张是不同的图片。

感知哈希算法的优势在于快捷简单，不会受到图片尺寸、比例大小的影响，但同样也存在一个严重的缺点，便是图片的内容不能更改，如果在原图上加上几个文字，图片搜索引擎就认不出来了，所以，感知哈希算法的最佳用途便是通过缩略图检索原图。实际运用中，大多使用的是更为强大的 pHash 算法和 SIFT 算法，它们可以识别出图像的变形，如果变形度不超过 25%，就可以找到原图。虽然这两个算法相对复杂，但运行原理与上面的感知哈希算法是一样的，就是将图片转化成 Hash 字符串再进行比较。

(二) 实验内容设计

选择目前国内规模最大、使用人数最多的三大图片搜索引擎——百度、360、搜狗，分别对其进行以图搜图的检索效率测试，然后探究以图搜图功能在舆论事件中对舆论的实时监控和导向作用，最后对比图片搜索与文字搜索的优劣势。

(三) 实验方法

1) 以图搜图追踪单一图片源头并测试其搜索效率

从网页和搜索引擎图片库中随机选择样本图片，将样本图片上传至图片搜索引擎，通过其以图搜图功能对样本图片进行识别，检索样本图片出现的网站源头，每个图片搜索引擎重复试验 400 次(连续 4 天在同一时段，100 次/天)，记录实验数据和结果。

2) 比较搜索引擎检索结果、汇总门户网站

通过微博、腾讯新闻等软件关注热点舆论事件的进程，及时获取与事件相关度高的图片作为样本。对样本图片每日的检索结果进行记录；通过比较总检索门户网站数，体现各搜索引擎的检索资源的广度与时效性；通过比较样本图片被不同搜索引擎检索到的重复次数，筛选信息传播较广的门户网站。

3) 数据整理分析方法

对各网址进行归一化处理，使之系统化，制作图表进行展示，使结果更直观，以方便比较。先对结果进行预处理，数据的预处理是数据整理的先前步骤，它是在对数据分类或分组之前所做的必要处理，内容包括网址的审核、排序等[1]。首先是将网址进行简化，省去前缀“http”与后缀“/”，并利用表格功能进行排序，对同一门户网站的网址进行归类，并得到真实的检索所得门户网站数。再制作统计表，统计表是显示统计数据的方式。检索结果经过预处理后，可进一步做整理、显示，制作成表格、图表进行展示。最后通过比较各图片搜索引擎以及汇总门户网站，结合具体的检索内容与相关结果，探究如何运用图片搜索引擎实现精准地舆论实时监控和方向把控。

3. 实验与分析

(一) 对图片搜索引擎“以图搜图”检索效率的实验与分析

从网页或搜索引擎图片库中随机选择的图片。

1) 百度图片搜索引擎

百度图片搜索引擎“以图搜图”称为百度识图，于2010年12月13日正式上线，百度是中国互联网产品中第一个提供以图搜图功能的搜索引擎。百度识图通过对图像底部局部特征的对比，并能根据图片库中存在的相似图片资源猜测用户上传图片对应的文本内容。从而满足用户寻找图片资源、模糊图换高清图、真伪识别等需求。

百度图片搜索引擎“以图搜图”的效率测试先后进行400次，读取样本图片进行源头追踪，记录实验次数与结果，所有实验数据绘制成为表1，对表1的数据进行统计与计算绘制成为表2。

Table 1. Efficiency test of Baidu picture search engine “Graph Search”

表 1. 百度图片搜索引擎“以图搜图”的效率测试

日期	实验次数	成功次数	成功率
9.4	100	73	0.7300
9.5	100	69	0.6900
9.6	100	81	0.8100
9.8	100	66	0.6600

Table 2. The test results of Baidu picture search engine “Graph Search”

表 2. 百度图片搜索引擎“以图搜图”测试结果

总实验次数	成功次数	成功率
400	289	0.7225

2) 360 图片搜索引擎

相比较百度，360 图片搜索引擎的界面更为干净，它为用户解决了两大问题，首先是搜索结果不掺杂广告信息，直接展示用户所需要的图片内容；其次是，搜索结果屏蔽了大量利用图片诱惑网民进入的木马网站，更为安全高效。

360 图片搜索引擎“以图搜图”的效率测试先后进行400次，读取样本图片进行源头追踪，记录实验次数与结果，所有实验数据绘制成为表3，对表3的数据进行统计与计算绘制成为表4。

Table 3. Efficiency test of 360 picture search engine “Graph Search”

表 3. 360 图片搜索引擎“以图搜图”的效率测试

日期	实验次数	成功次数	成功率
9.11	100	70	0.7000
9.12	100	63	0.6300
9.13	100	52	0.5200
9.14	100	58	0.5800

Table 4. Test result of 360 picture search engine “Graph Search”

表 4. 360 图片搜索引擎“以图搜图”测试结果

总实验次数	成功次数	成功率
400	243	0.6075

3) 搜狗图片搜索引擎

搜狗图片搜索引擎在以图搜图的方式上进行了更为快捷简便的设计，除了与百度和360一样，通过

上传图片、鼠标拖拽、输入图片 URL 地址等方式进行以图搜图以外，搜狗还可以通过鼠标在图片上悬停便能够完成搜索[2]，实现浏览器与搜索引擎的有机融合，大幅度提升了用户图片搜索的效率。

搜狗图片搜索引擎“以图搜图”的效率测试先后进行 400 次，读取样本图片进行源头追踪，记录试验次数与结果，所有实验数据绘制成为表 5，对表 5 的数据进行统计与计算绘制成为表 6。

Table 5. Efficiency test of Sogou picture search engine “Graph Search”

表 5. 搜狗图片搜索引擎“以图搜图”的效率测试

日期	实验次数	成功次数	成功率
9.11	100	60	0.6000
9.12	100	88	0.8800
9.13	100	76	0.7600
9.15	100	96	0.9600

Table 6. Test result of Sogou picture search engine “Graph Search”

表 6. 搜狗图片搜索引擎“以图搜图”测试结果

总实验次数	成功次数	成功率
400	320	0.8000

将表 2，表 4，表 6 的实验数据结果进行汇总，绘制成图 1，将百度，搜狗和 360 三大图片搜索引擎的实验成功率进行直观表现。

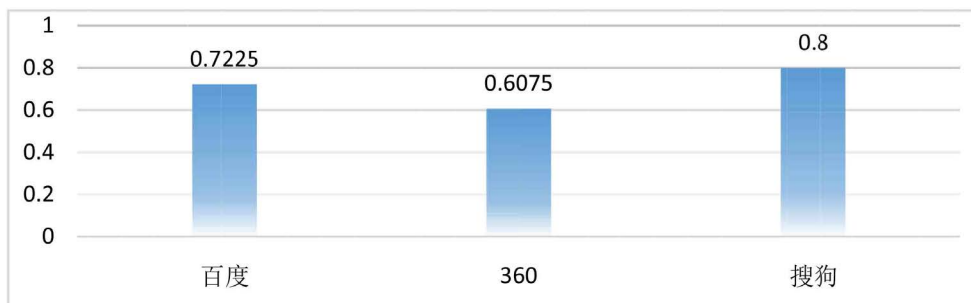


Figure 1. Baidu, Sogou, 360 “Graph Search” success rate contrast map

图 1. 百度，搜狗，360 “以图搜图”成功率对比图

4) 实验结果分析

从图 1 中的图像数据可以直观体现出三大图片搜索引擎“以图搜图”检索效率的差异。搜狗 80%的成功率为最高，紧随其后的是百度 72.25%，360 的 60.75%成功率为最低。考虑到以图搜图的核心技术是感知哈希算法，也就存在不可避免的缺陷，如果读取图片加有过多的文字或者图片本身自带较多的文字，搜索引擎便无法进行相似检索，从而对以图搜图的搜索效率产生影响，所以三者的成功率均已达到较高比例。

“以图搜图”功能凭借着较高的成功率实现对单一图片源头的检索追踪，可以为公安机关运用图片搜索引擎进行舆情导控提供基础。就图片搜索引擎“以图搜图”的搜索效率和实际操作的便捷程度而言，搜狗无疑是首选；但是就使用范围和搜索信息范围的广度而言，拥有中国最大图片资源库的百度则是首选；而就图片搜索的体验感和安全性而言，360 则要优于百度和搜狗。综合考虑，公安机关在追查网络图片源头时，可以根据需要，考虑使用百度、搜狗和 360 图片搜索引擎。

(二) 重庆公交车坠江事故事件的检索实验与分析

1) 事件概述

2018年10月28日10时08分,在重庆市万州区万州长江二桥上,一辆22路公交车突然越过大桥中心实线与迎面驶来的一辆红色小轿车碰撞后坠入江中。

2018年10月29日14时20分,公安机关通过走访调查和接警中心的报警记录,初步确认15人失联(1名司机,14名乘客)。

2018年10月31日0时50分,救援部队打捞出公交车的黑匣子并移交当地公安机关。

截止2018年11月1日15时,救援部队已经打捞出13具遗体,并确认身份,仍有2人失联。

2018年11月2日,公安机关公布公交车坠江原因,根据公交车内部黑匣子监控显示,由于一名女乘客与司机发生口角,互殴导致车辆失控坠江。

2) 舆情发展过程

事发当天,多家媒体报道了此事件,无一例外地指出是公交车为了避让一辆逆行的红色轿车才导致了悲剧的发生,舆论的声讨直指轿车女司机;10月29日,一段事发地车辆的行车记录视频被公布后,舆论第一次转向,大家纷纷同情女司机,而公交车司机成为了众矢之的;11月2日,公交车内部监控视频公布,真相浮出水面,舆论形势达到了顶峰,同时舆论第二次转向,抢方向盘的女乘客被集体声讨。

3) 检索图片样本选择

筛选与事故相关度较高的图片作为样本。利用图片搜索引擎对它们进行检索。见下图:图2警情通报(Police intelligence Bulletin),图3事故现场1(Accident scene 1),图4事故现场2(Accident scene 2),图5事故现场3(Accident scene 3),图6公交监控(Bus interior monitoring),图7打捞现场1(Salvage site 1),图8打捞现场2(Salvage site 2),图9救援现场(Rescue site)。

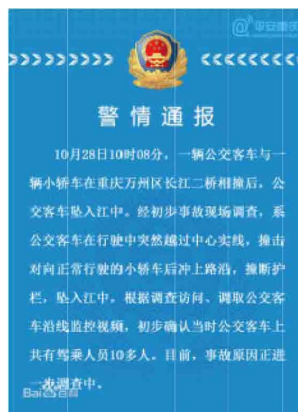


Figure 2. Police intelligence bulletin
图2. 警情通报



Figure 3. Accident scene 1
图3. 事故现场1



Figure 4. Accident scene 2
图 4. 事故现场 2

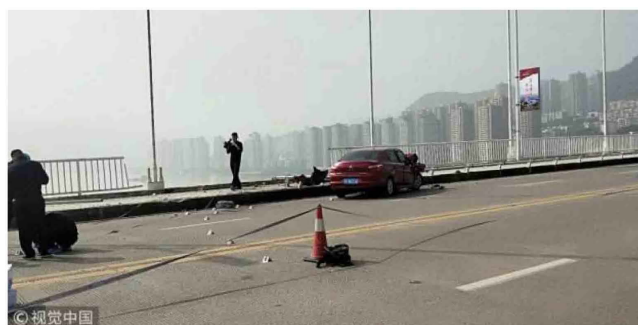


Figure 5. Accident scene 3
图 5. 事故现场 3



Figure 6. Bus interior monitoring
图 6. 公交监控

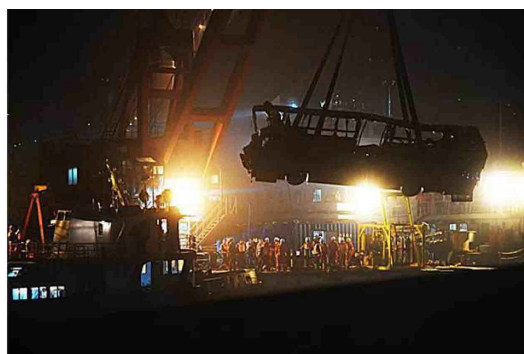


Figure 7. Salvage site 1
图 7. 打捞现场 1



Figure 8. Salvage site 2

图 8. 打捞现场 2



Figure 9. Rescue site

图 9. 救援现场

4) 图片搜索引擎“以图搜图”的检索结果

将获取的图 1~图 8 分别上传至百度、360、搜狗各个图片搜索引擎进行检索，得到的每日检索的门户网站总数，制作成表 7，为了进行更为直观的比较，制成表 8、图 10。

然后以门户网站为量度统计，通过三大图片搜索引擎对样本图片进行检索，挑选出重复次数较高的门户网站制作成表 9。

Table 7. Daily search portal statistics table (data represents search engine search portal number)

表 7. 每日检索门户网站数统计表(数据表示搜索引擎检索门户网站数)

日期	百度	360	搜狗
10.28	1	1	3
10.29	4	2	11
10.30	36	4	15
10.31	49	5	22
11.1	69	9	24
11.2	113	25	41

Table 8. Statistical table of total search results of portal websites (time 2018-10-28~2018-11-2)

表 8. 门户网站检索结果总数统计表(时间 2018-10-28 至 2018-11-2)

搜索引擎	百度	360	搜狗
检索结果总数	113	25	41

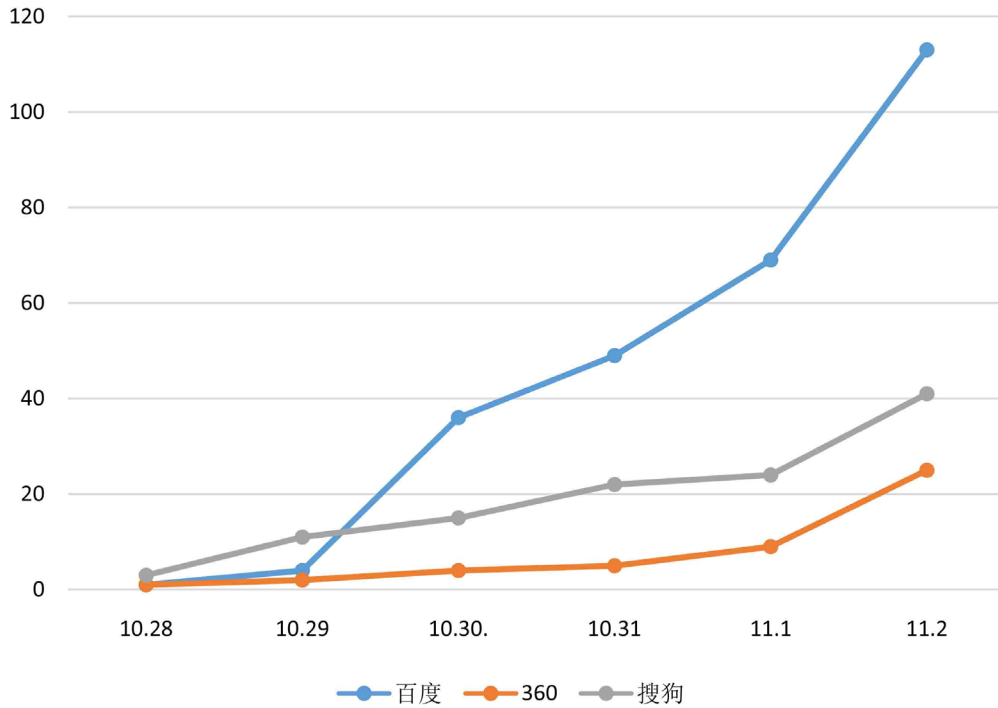


Figure 10. Daily search portal statistics table (broken line chart)
图 10. 每日检索门户网站数统计表(折线图)

Table 9. Daily repeated common website statistics table (data in the table indicate the number of times the portal site was retrieved repeatedly on that day)

表 9. 每日重复共有网站统计表(表中数据表示当日该门户网站被重复检索的次数)

门户网站	10.28	10.29	10.30	10.31	11.1	11.2	总计
Baijiahao.baidu.com	0	2	26	11	11	7	57
Ifeng.com	2	0	1	1	3	1	8
www.xinhuanet.com	0	1	3	1	5	1	11
www.youth.cn	0	0	1	0	1	0	2
www.CYOL.com	0	0	0	1	1	0	2
Weibo.com	1	3	1	2	0	5	12
sina.com.cn	0	0	1	1	1	1	4
news.163.com	0	1	1	0	2	0	4
news.k618.cn	0	0	0	0	2	0	2
Chinanew.com	0	0	0	1	0	2	3
www.haiwainet.com	0	0	1	0	1	1	3
www.thepaper.cn	0	1	0	1	1	2	5

5) 舆情分析

10月28日当天,有关私家车逆行导致公交车坠江的消息已经在网络上传播开,许多媒体和网络大V未经证实地转载报道,引发众多网民对女司机的声讨。10月29日,舆论又将矛头指向公交车司机。直到11月2日警方公布事发的真正原因,在这期间,舆论对于事故发生原因的讨论一直呈现上升趋势。反观事发地警方,并没能第一时间有效地引导舆论,控制谣言的传播。

首先,警方可以在事件发生的第一时间对网络舆情进行实时监控,筛选出未经证实的虚假报道。这时,图片搜索引擎的优势就凸显出来了,选取出传播量大,与现场高度相关的图片,利用以图搜图功能追踪锁定图片源头,筛选出未经证实的内容,依法追究负面网络舆论传播者的法律责任,将虚假信息扼杀在摇篮,保证网络信息的真实性[3]。

其次,在利用图片搜索引擎排除虚假信息的同时,警方还可以把控社会舆论的方向,对其进行正确的引导。然而在此次事件中,在事发地警方11月2日的事故原因通气会之前,大多都是发布事故调查的通报,这是远远不够的。对于发布虚假信息的网站或个人,警方应当及时处理并联系权威媒体、网站对处理结果和事故调查的进度做实时的报道,避免大众被虚假信息蒙蔽双眼,保证网络舆论的完整性,正确性,及时性[4]。

(三) 图片搜索与文字搜索比较

将图1至图8分别上传至百度、360、搜狗图片搜索引擎进行检索,对图片检索结果进行统计记录;提取关键词“重庆长江大桥”,“公交车坠江”在百度、360、搜狗搜索引擎进行检索,对文字检索结果进行统计记录(表10)。

Table 10. Statistical table of the search result

表 10. 检索结果统计表

搜索引擎	百度	360	搜狗
图片检索结果	113	25	41
文字检索结果	455,000	32,614	3876

单从检索获取的资源量上看,文字检索占据绝对优势。但是文字检索存在着准确度较低的问题,大量检索结果与舆论事件并不相关,就该案件而言,会检索到其他发生在公交车上的事件。虽然文字检索可以框定用户想要的信息的时间区间,从而缩小信息量的范围,但其检索所得信息的准确度依旧不高。另外,有些案件提取的关键词较为宽泛,甚至有些案件无法提取关键词,这就极大限制了文字检索的使用范围。

与文字检索相比,图片检索的准确性是有明显优势的,并且操作方式更为便捷,但是图片搜索引擎本身的工作原理也造成了它检索的局限性。以图搜图是基于图片的搜索,所以检索结果都是包含样本图片的内容,而那些与舆论事件相关的全文字报道则不在检索结果的行列,这就导致了图片检索在信息量的广度上输于文字检索。

总而言之,图片检索拥有较高的准确性,值得信任,但所得的网站信息数量较少是其不可避免的局限性。

4. 结语

作为新媒体时代的新兴力量,图片搜索引擎的以图搜图功能有着巨大的发展空间,并且正在朝实用化的方向逐步迈进。但现阶段,图片搜索引擎在公安工作中的实战化运用还不成熟,存在一些观念或者认识上的问题,使图片搜索引擎无法高效地运用到舆情导控工作中[5]。未来的图片搜索引擎实战化运用,需要形成关于网络舆论的独立的、系统的法律体系,为公安机关运用图片搜索引擎进行舆情导控提供坚实的法律保障;需要建立舆情导控专项工作平台,使图片搜索引擎的运用系统化、高效化;需要推动图片搜索引擎的运用由试点到普及的进程,为公安机关处理网络舆论事件提供更为有效的手段。相信在未来的公安工作中,图片搜索引擎将会更广泛地运用于舆情导控工作,并发挥重大作用。

参考文献

- [1] 季永春, 朱赫, 张倩兰. 基于计算机的统计学原理分析[J]. 江西建材, 2017(2): 284.
- [2] 周长建. 基于多示例学习的视频字幕提取算法研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工程大学, 2012.
- [3] 杨芳. 坚持新闻真实性[J]. 西部广播电视, 2016(21): 28.
- [4] 郑小兰. 公共关系理论与实务[M]. 上海: 立信会计出版社, 2006: 11.
- [5] 钱斌. 基于语义网的舆情监控系统的设计与实现[D]: [硕士学位论文]. 兰州: 兰州大学, 2016.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-6753, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: jisp@hanspub.org