

DOI: 10. 12138/j. issn. 1671-9638. 20204174

· 论 著 ·

## 中国医院感染控制电子信息服务系统热点分析 ——基于 SIFIC 论坛的数据挖掘

曹煜隆<sup>1</sup>, 单 娇<sup>2</sup>, 包小源<sup>3</sup>, 金 梦<sup>3</sup>, 周 璐<sup>4</sup>, 张 超<sup>5</sup>, 孙庆芬<sup>6</sup>, 韩玲样<sup>7</sup>, 高 燕<sup>1</sup>, 胡必杰<sup>8</sup>

(1. 北京大学人民医院医院感染管理办公室, 北京 100044; 2. 北京积水潭医院疾病预防控制处, 北京 100035; 3. 北京大学医学信息学中心, 北京 100191; 4. 北京中医药大学中医学院, 北京 100029; 5. 北京左医科技有限公司, 北京 100044; 6. 赤峰学院附属医院感染管理科, 内蒙古 赤峰 024000; 7. 铜川市妇幼保健院医院感染管理科, 陕西 铜川 727007; 8. 复旦大学附属中山医院感染管理科, 上海 200032)

**[摘要]** **目的** 回顾近十年来我国医院感染控制电子信息服务系统的热点, 总结医院感染管理人员关注热点的趋势变化。**方法** 基于 Python 爬虫程序设计方法, 抓取 2007 年 4 月 6 日—2017 年 12 月 31 日在上海国际医院感染控制论坛(SIFIC)中发表的内容, 对其进行中文分词处理并提取出现频次位于前 5 位的有效词作为热点词汇进行分析。**结果** 共获取 2007—2017 年 SIFIC 论坛 93 个板块共 3 112 840 条数据。从年度纵向分析来看, 年度热点词汇前 3 位主要为“消毒”、“监测”、“培训”等。2009 年全球暴发新甲型 H1N1 流感, “甲型流感”成为年度第一热词; 2017 年 10 个医院感染管理新规范正式施行, “规范”成为年度第二热词。从重点版块横向来看, 每一专业版块均有自己不同的侧重方向, 表现出医院感染管理学科的复杂性与多样性。**结论** 通过对医院感染控制领域电子信息服务系统的热点分析, 可以了解该领域在实际工作中关注重点的变化趋势, 从而进一步提高对热点问题的关注度, 不断提高医院感染控制能力。

**[关键词]** 医院感染; 电子信息服务系统; 论坛; Python 语言

**[中图分类号]** R197.323

## Hot spots of healthcare-associated infection control electronic information service system in China——Data based on SIFIC BBS

CAO Yu-long<sup>1</sup>, SHAN Jiao<sup>2</sup>, BAO Xiao-yuan<sup>3</sup>, JIN Meng<sup>3</sup>, ZHOU Lu<sup>4</sup>, ZHANG Chao<sup>5</sup>, SUN Qing-fen<sup>6</sup>, HAN Ling-yang<sup>7</sup>, GAO Yan<sup>1</sup>, HU Bi-jie<sup>8</sup> (1. Office of Healthcare-associated Infection Management, Peking University People's Hospital, Beijing 100044, China; 2. Department of Disease Control, Beijing Jishuitan Hospital, Beijing 100035, China; 3. Medical Information Center, Peking University, Beijing 100191, China; 4. School of Traditional Chinese Medicine, Beijing University of Traditional Chinese Medicine, Beijing 100029, China; 5. Beijing Zuoyi Technology Co. Ltd., Beijing 100044, China; 6. Department of Healthcare-associated Infection Management, Affiliated Hospital of Chifeng University, Chifeng 024000, China; 7. Department of Healthcare-associated Infection Management, Tongchuan City Maternal and Child Health Care Hospital, Tongchuan 727007, China; 8. Department of Healthcare-associated Infection Management, Zhongshan Hospital, Fudan University, Shanghai 200032, China)

**[Abstract]** **Objective** To review the hot spots in healthcare-associated infection(HAI) control electronic information service system in China in recent 10 years, and summarize the trend changes of hot spots of HAI management

**[收稿日期]** 2019-08-22

**[作者简介]** 曹煜隆(1991-), 男(汉族), 北京市人, 助理研究员, 主要从事医院感染管理研究。

**[通信作者]** 高燕 E-mail: caoyulongchn@163.com

personnel. **Methods** Based on the python scrapy programming method, this paper captured the content published in the Shanghai International Forum for Infection Control(SIFIC) from April 6, 2007 to December 31, 2017, processed Chinese word segment, and extracted the top 5 available words as hot words for analysis. **Results** A total of 3 112 840 data from 93 sections of SIFIC from 2007 to 2017 were obtained. According to the annual longitudinal analysis, the top 3 hot words were “disinfection”, “monitoring”, and “training”. In 2009, the global outbreak of novel influenza A (H1N1), “influenza A” became the first hot spot word of the year; in 2017, 10 new regulations for HAI management were officially implemented, and “regulation” became the second hot spot word of the year. From horizontal perspective of key sections, each professional section has its own different focus direction, showing the complexity and diversity of HAI management discipline. **Conclusion** Through the hot spot analysis on the electronic information service system in the field of HAI control, changing trend of focus in the field of practical work can be understood, the focus on hot issues can be further improved and HAI control ability can be improved.

[**Key words**] healthcare-associated infection; electronic information service system; forum; Python

近年来,互联网技术的进步使人们的工作方式逐渐发生改变,人们通过网络平台更便捷的学习到新的知识,发表自己的观点。在医院感染管理领域中,上海国际医院感染控制论坛(Shanghai International Forum for Infection Control, SIFIC)就是一个非常突出的例子,该论坛成立十余年来,越来越多的医院感染控制专(兼)职人员参与到论坛中,与他人互动、分享各种内容。SIFIC 论坛给全国医院感染管理人员提供了高效的知识共享和交流平台,同时也吸引了大量的医院感染管理人员参与到相关内容的研究。网络爬虫被广泛用于互联网搜索引擎和其他类似的网站,通过代码可以让程序自动采集所有能够访问到的页面内容,最终获取或更新这些网站的内容和检索方式。近年来人们利用 Python 对 Twitter、Facebook 等知名社交平台展开了一系列的分析<sup>[1-3]</sup>,但是目前尚无利用 Python 技术对医学相关网站及论坛的研究,因为对不熟悉爬虫技术的医学研究人员而言这无疑是一个极大的挑战。本文基于 Python 语言的 scrapy(爬虫框架)获取 SIFIC 论坛的各板块数据,并进行分析,以期回顾近十年来我国医院感染控制电子信息服务系统的热点,总结医院感染管理人员关注热点的趋势变化,从而进一步提高对热点问题的关注度,不断提高医院感染控制能力。

## 1 资料与方法

1.1 数据来源 选取 SIFIC 论坛 2007 年 4 月 6 日—2017 年 12 月 31 日的数据作为此次数据来源。SIFIC 论坛于 2007 年 4 月 6 日创建,论坛会员遍及国内几十多个省市,多个专业;现注册会员近 13 万人,专业版块 155 个,是目前中国规模最大、原创资料最

多、专家队伍最强的医院感染预防与控制论坛。

1.2 数据获取 通过 Python 语言的 scrapy 获取主题网络,爬虫从论坛某一个页面开始,读取标题内容,找到在网页中的其他链接地址,然后通过链接地址寻找下一个网页,这样一直循环将 2007 年 4 月 6 日—2017 年 12 月 31 日 SIFIC 论坛上所有的网页标题进行抓取,将所提取的信息组合成 DataFrame 格式,自此完成所有数据的采集工作。

1.3 中文分词及词频统计 为进行热点词的统计与分析,本研究使用基于 Python Jieba (Chinese for “to stutter”, 0.38 版本)软件包对爬取的发帖标题文本串进行分词,分词时利用停用表对诸如“的、吗、和”非实意词进行剔除,最后利用 Python 的 Collections. Counter 功能对分词结果进行词频统计,并在此基础上进行综合分析。

## 2 结果

2.1 基本概况 共爬取 SIFIC 论坛 93 个板块共 3 112 840 条数据,其中主题帖 155 576 条,平均每个主题回帖 19.01 次,平均每日新增 794 帖;累计注册会员共 151 471 名,其中管理员和超级版主 131 名,平均每人发帖 20.55 条。

2.2 年度热点分析 按照发帖的时间逐年筛选每一年中出现次数位于前五的高频词频进行分析,见表 1。从年度发帖数量上看,2007—2013 年呈逐年递增的趋势,2013 年后论坛发帖数量有所下降,这与 2013 年底 SIFIC 论坛筹备开通官方微信有着极为密切的关系。从年度高频热词来看,“消毒”、“监测”、“培训”、“手术”、“手卫生”、“耐药”等词是医院感染管理人员关注的永恒主题。2009 年全球暴发新甲型 H1N1 流感,“甲型流感”成为年度第一热

词;2017 年 10 个医院感染管理新规范正式施行,“规范”成为年度第二热词。

2.3 重点版块热点分析 在排除类似于“新手集结号”、“商城”、“休闲时光”等与医院感染管理工作不相关的版块和发帖数较少的冷门版块后,对关注较多的 8 大重点热门版块中的 31 个子版块进行高频热词分析。对于每一版块的热词筛选,排除该版块或子版块的题目词及相关词汇,如“重点部位—血流感染 BSI”版块,高频词汇则排除“中心”、“静脉”、“导管”、“血流”、“感染”及“BSI”等词,仅对有可能表现版块关注方向的词进行归纳整理,并合并类似于“洗手液”、“手消毒剂”等词义相近的词汇。从重点版块横向来看,每一专业版块均有自己不同的侧重方向,表现出医院感染管理学科的复杂性与多样性,见表 2。

**表 1** 2007—2017 年 SIFIC 论坛年度前 5 位高频热词分析  
**Table 1** Annual top 5 high frequency words in SIFIC in 2007 - 2017

年份	总帖数	热点词汇				
		第 1 位	第 2 位	第 3 位	第 4 位	第 5 位
2007	3 594	消毒	监测	药物	标准预防	微生物
2008	6 540	消毒	监测	抗菌药物	预防	耐药
2009	11 391	甲型流感	消毒	卫生部	监测	标准
2010	15 582	消毒	监测	手术	灭菌	预防
2011	22 338	消毒	监测	手术	抗菌药物	培训
2012	33 909	消毒	监测	手术	培训	耐药
2013	37 466	消毒	监测	培训	手术	耐药
2014	27 291	消毒	监测	培训	手术	耐药
2015	26 290	消毒	监测	培训	手卫生	手术
2016	26 825	消毒	监测	培训	手卫生	手术
2017	28 101	消毒	规范	监测	培训	医疗废物

**表 2** 2007—2017 年 SIFIC 论坛版块前 5 位高频热词分析  
**Table 2** Top 5 high frequency words in each section of SIFIC in 2007 - 2017

版块	子版块	总帖数	热点词汇				
			第 1 位	第 2 位	第 3 位	第 4 位	第 5 位
基础感控	手卫生	13 992	洗手	医疗废物	依从性	洗手液	宣传
	清洁消毒灭菌技术	16 496	监测	紫外线	空气消毒	灭菌器	消毒剂
	职业安全	6 140	防护	处理	口罩	隔离	针刺伤
	医疗废物 污水管理	5 529	处置	分类	一次性	暂存	锐器
	一次性医疗器械/器具和消毒药械管理	3 822	审核	无菌	产品证件	灭菌	有效期
重点部位	手术部位感染 SSI	5 130	切口	监测	术后预防	一类切口	剖宫产
	血流感染 BSI	946	培养	指南	监测	诊断	操作
	医院内肺炎 HAP/呼吸机相关肺炎 VAP	1 146	诊断	消毒	监测	指南	湿化器
	尿路感染 UTI	843	更换	诊断	监测	治疗	指南
重点部门	ICU	2 514	目标性监测	消毒	呼吸机	层流	耐药
	新生儿室与 NICU	2 042	消毒	暖箱	监测	奶瓶	标准
	消毒供应中心	12 613	器械清洗	监测	包装管理	灭菌器	流程
	手术室	7 080	洁净	消毒	层流	空气监测	灭菌
	产房与人流室	1 290	消毒	分娩	胎盘	隔离	布局
	血液透析室	3 969	监测	消毒	透析液	内毒素	丙肝
	口腔科	2 082	清洗消毒	手机	规范	监测	灭菌器
	内镜室	3 156	清洗消毒	灭菌	胃镜	监测	戊二醛
临床微生物与多重耐药菌	临床微生物标本采集与检测	3 314	送检率	耐药	血标本	痰标本	药敏结果
	多重耐药菌监测、防控与诊治	7 274	隔离	MRSA	Ab	培训	定植

续表 2 (Table 2, Continued)

版块	子版块	总帖数	热词词汇				
			第 1 位	第 2 位	第 3 位	第 4 位	第 5 位
抗菌药物应用与管理	抗菌药物基础理论与规范指南	5 845	切口	合理用药	耐药	围手术期	送检
感染病	感染病诊治	4 499	是否属于院感	标准	发热	感染部位	培养
	常见传染病诊疗与防控	6 670	流感	H7N9 禽流感	H1N1 甲型	艾滋病	手足口
	我国法定传染病管理	2 950	报告	门诊	消毒隔离	梅毒	防控
	感染暴发事件及医院感染纠纷案例	3 518	演练	应急处置	报告	预案	地震
医院感染监测	医院感染综合性监测	3 735	现患率	漏报	感染率	手术	指标
	医院感染目标性监测	2 764	手术切口	ICU	调查方案	耐药	现患率
	医院环境微生物及消毒灭菌效果监测	5 393	采样	空气监测	灭菌	培养	紫外线
管理工具及科研	管理工具的理论与实践	3 482	质量	制度	临床	品管圈	院感委员会
	文献、论文、科研、基金	1 443	杂志	英语	SCI	课题	统计
	Meta 分析与系统评价	278	循证	培训班	论文写作	软件	文献
	感染控制与医院流行病学	670	培训班	讲评	临床实践	文献	暴发调查

MRSA:耐甲氧西林金黄色葡萄球菌(methicillin-resistant *Staphylococcus aureus*, MRSA); Ab:鲍曼不动杆菌(*Acinetobacter baumannii*, Ab)

### 3 讨论

医院感染与医院的建立相依并存,是当前公共卫生领域的一个重要问题,并直接影响着医疗质量和患者的安全。医院感染管理是一项复杂而庞大的系统工程,政策性、学术性及专业性均较强。专业方面既涉及卫生统计学和流行病学的知识,又涵盖临床医学、微生物学、免疫学、药学和护理学等诸多学科的知识。随着医学科学的进步与发展,尤其是日新月异的治疗方法的开展,抗菌药物的广泛应用,新的演变趋势使医院感染管理变得越来越复杂化与多样化<sup>[4]</sup>,对从事医院感染监控工作的专(兼)职人员知识更新的要求也变得越来越高的。目前,医院感染管理专业队伍建设较为薄弱<sup>[2]</sup>,从基层医疗机构到三级医院,医院感染管理人员队伍很不稳定,多以兼职人员为主,工作强度较大,故医院感染管理人员如何在紧张的工作中获取、更新最有价值的知识内容便显得尤为重要。

互联网的持续发展将信息爆炸时代的知识交互带入了一个全新的阶段,使得沟通摆脱了空间的束缚,利用论坛这一渠道进行即时信息的传递与交流,实现高效率协作,强调人与人相互连接。SIFIC 为中国医院感染监控工作人员搭建了良好的交流平台,使中国医院感染预防与控制水平与国际接轨,其会员来自医疗领域的各个部门与专业,拥有旺盛的人气、良好的交流氛围及广阔的交流空间。

以 2009 年为例,全球超过 207 个国家和地区报告了新甲型 H1N1 流感实验室确诊病例,总数超过 62 万例,包括 7 820 余例死亡病例,世界卫生组织将预警级别升至最高级别 6 级<sup>[5]</sup>,2009 年 SIFIC 论坛关注的热点也随之变为了“甲型流感”;而针对全球流感背景下,环境、物体表面如何进行清洁“消毒”;流感的“监测”与上报;医护人员如何在流感大暴发的环境中采取“标准”预防的措施与应对这种紧急情况医院感染专业人员采取什么样的强化“培训”则变为了该年关注频次位于前 5 位的词汇,较好的体现了论坛紧贴国际热点变化的时效性。

从重点版块横向来看,每一版块均有自己不同的侧重方向,表现出医院感染管理学科的复杂性与多样性。以“重点部位—手术部位感染 SSI”中的热词词汇为例,“切口类型”、“目标监测”、“术后预防 SSI”、“剖宫产”等为 SSI 关注的热点,关于手术后切口感染影响因素的研究较多,由于发生切口感染不仅给患者带来身体的痛苦,也会对其心理造成不良影响,对治疗失去信心,降低治疗的依从性,更严重者会出现血液系统感染疾病,甚至增加死亡风险<sup>[6-7]</sup>。所以,医院感染监控人员的关注点更多在如何预防手术切口感染的发生,对手术部位感染的切口类型和常发生手术部位感染的手术类型进行探讨分析是极为必要的。

本研究将大数据的理念成功应用在爬虫分析中,高效的识别出以往关注的热点,对今后的研究提供了参考性的建议。但同时本研究也存在一定的局

限性,由于数据所限,未能考虑使用和未使用本论坛的医院感染管理相关人员的差异,因此分析出的结果可能存在一定的偏倚。此外由于本研究纳入分析的数据量较大,涉及面较广,不能对每一版块的高频词汇逐一进行分析,仅为医院感染管理工作人员描绘近年来国内医院感染监控关注的方向,发现广泛关注点在哪里,为日后研究提供新思路,每一版块相对应的工作人员也可根据本研究所示的热点结果找寻自己的研究方向<sup>[8-9]</sup>。SIFIC论坛创办至今已十余年,标志着中国医院感染管理已进入科学化、信息化的阶段,本研究以实例说明爬虫数据挖掘的研究意义,今后可以以大数据为基础,探索更多的研究方向。

致谢:感谢上海国际医院感染控制论坛(SIFIC)提供相关数据。

#### [参考文献]

- [1] Carlisle JE, Patton RC. Is social media changing how we understand political engagement? An analysis of facebook and the 2008 presidential election[J]. Polit Res Quart, 2013, 66(4): 883 - 895.
- [2] Strafling N, Kramer NC. Learning together on Facebook et al. The influence of social aspects and personality on the usage of social media for study related exchange[J]. Gruppendynamik und Organisationsberatung, 2013, 44(4): 409 - 428.
- [3] Sanner MF. Python: a programming language for software in-

tegration and development[J]. J Mol Graph Model, 1999, 17(1): 57 - 61.

- [4] 李丽华. 新形势下如何做好医院感染管理工作[J]. 中国医药指南, 2013, 11(32):585 - 586.
- [5] 任晓旭,宋国维. 2009年新甲型H1N1流感[J]. 中国小儿急救医学, 2009, 16(6):528 - 532.
- [6] 张麦玲,李艳华,朱小芳,等. 外科手术患者发生切口感染的手术室相关因素分析及防治策略[J]. 中华医院感染学杂志, 2018, 28(1):152 - 155.
- [7] 戴浩,戴佩芬,毛斌存,等. 外科手术患者术后切口感染相关因素分析[J]. 中华医院感染学杂志, 2018, 28(20):3101 - 3103.
- [8] 王莹,李源,金学兰,等. 基于共现分析的国际医院感染研究热点可视化探析[J]. 中华医院感染学杂志, 2017, 27(18): 4312 - 4316.
- [9] 江建忠,贾素玉,宋倩. 国际医院感染控制研究热点分析[J]. 中华医院感染学杂志, 2016, 26(24):5733 - 5736.

(本文编辑:陈玉华)

**本文引用格式:**曹煜隆,单娇,包小源,等. 中国医院感染控制电子信息服务系统热点分析——基于SIFIC论坛的数据挖掘[J]. 中国感染控制杂志, 2020, 19(2): 164 - 168. DOI: 10. 12138/j. issn. 1671 - 9638. 20204174.

**Cite this article as:** CAO Yu-long, SHAN Jiao, BAO Xiao-yuan, et al. Hot spots of healthcare-associated infection control electronic information service system in China——Data based on SIFIC BBS[J]. Chin J Infect Control, 2020, 19(2): 164 - 168. DOI: 10. 12138/j. issn. 1671 - 9638. 20204174.