

科技论文 10 个无效关键词计量学分析*

伍锦花 陈灿华

《中南大学学报(自然科学版)》编辑部,410083,长沙

摘要 科技论文中存在大量不能反映论文主题的无效关键词,降低了论文显示度和学术影响力。本文以中国知网为文献检索数据库,针对科技论文中“应用”“对策”“问题”“管理”“设计”“措施”“影响”“现状”“研究”“发展”这 10 个无效关键词展开计量学分析,发现 2015—2019 年这 10 个无效关键词的出现频率相对稳定,其共现关键词中无效关键词的比例非常高;“学院学报”及硕士论文中常见无效关键词的累计词频明显比“大学学报”以及博士论文的高,说明前者的关键词标引质量较差。科技论文作者及审读编校人员应重视关键词标引工作,避免因使用无效关键词而降低文献传播效果。

关键词 科技论文;无效关键词;计量学分析;关键词共现;文献检索;词频;学术影响力

Metrological analysis of common invalid keywords in scientific papers/WU Jinhua, CHEN Canhua

Abstract There exists a great amount of keywords that cannot reflect the theme accurately in scientific papers, which reduces the accessibility and academic influence of the scientific papers. In this paper, CNKI was used as the document retrieval database. Metrological analysis of 10 common invalid keywords in scientific papers was carried out, including “application” “countermeasure” “issue” “management” “design” “measure” “influence” “present situation” “research” and “development”. We found that occurrence of the above 10 common invalid keywords in scientific papers is relatively stable, and the proportion of invalid keywords is remarkably high in the co-occurrence network of these 10 keywords from 2015 to 2019. The accumulative word frequencies in scientific papers published by college journals and Master’s thesis are obviously higher than that in scientific papers published by university journals and doctoral dissertations in the past 5 years, which indicates that the indexing quality of the former is worse than that of the latter. We suggest researchers and editors pay more attention to the indexing of keywords, so as to avoid unsatisfactory viewed times of scientific papers due to usage of invalid key words.

Keywords scientific papers; invalid keywords; metrological analysis; co-occurrence of keywords; document retrieval; word frequency; academic influence

Authors’ address Editorial Department of Journal of Central South University (Science and Technology), 410083, Changsha, China

DOI:10.16811/j.cnki.1001-4314.2020.04.011

近年来,中国学者每年发表的科技论文有数百万

篇,如何在浩瀚的论文海洋中脱颖而出,吸引读者的注意力,提高论文传播力和显示度,关键词起着十分重要的作用^[1]。关键词是能够明确表达学术论文的主体研究内容或中心思想的词或词组^[2],其最初是为了满足计算机系统编制各种文献索引的需要而产生。作为文献检索的重要依据,关键词应具有专指性、全面性、规范性等特点。但实际情况是,科技论文中存在大量的含义宽泛、无实质内容的无效关键词^[3],如“应用”“分析”“问题”“研究”等等,大大降低了科技论文的检索效率和传播效果。此前,已有学者对无效关键词展开研究,但其主要针对高校社科学报论文以及农业科技类论文^[3-4],针对更广泛领域内的科技论文中常见无效关键词的计量学研究较少。为此,本文以中国知网为文献来源数据库,结合文献^[3-5]及编校实践经验,选取 10 个科技论文中出现频率较高、指向性不强、含义空泛的关键词,分别检索其在 2015—2019 年发表的科技论文中的出现频率,分析其年度分布规律及关键词共现网络。随后分别限定“大学学报”和“学院学报”以及“博士论文”和“硕士论文”(即博士和硕士的学位论文)为文献来源,对比分析常见无效关键词在不同类型期刊以及不同学历作者所著科技论文中的词频分布。最后分析常见无效关键词产生的原因及应对措施,以期提高科技论文作者和审读编校人员对关键词标引质量的认识。

1 检索范围和方法

检索范围:文献来源限定为中国知网“基础科学”“工程科技 I 辑”“工程科技 II 辑”“农业科技”“医药卫生科技”“信息科技”6 类目录下的期刊、会议、学位论文等(即本文所指科技论文)，“大学学报”和“学院学报”，“博士论文”和“硕士论文”；文献语种限定为中文；关键词限定为“应用”“对策”“问题”“管理”“设计”“措施”“影响”“现状”“研究”“发展”；发表时间限定为 2015-01-01—2019-12-31。

检索方法:以“关键词+发表时间”“关键词+发表时间+文献来源”为组合检索条件,在中国知网 6 大类科技文献目录下进行检索,获得各关键词在不同检索条件下的词频分布。为确保检索结果的有效性,每检索一次均随机选取若干篇论文,并核对这些论文

* 中国科技期刊卓越行动计划(C-194)

中是否包含目标关键词。最后,将所得结果输入 Excel 软件进行统计分析(检索时间为 2020-03-07—03-09)。

2 无效关键词的定义和分类

2.1 无效关键词的定义

无效关键词是指词义宽泛、不能准确传达论文主题内容、不利于检索的词或词组,或是未经公认的缩写词等^[3,5-6]。它们与论文主题相关性不大,不能反映单篇科技论文的特点,不具检索价值。

2.2 无效关键词的分类

无效关键词的分类目前尚无统一标准。本文根据文献^[3,6]将无效关键词分为以下3类:1)表示国别或地域的泛而不专的名词,如“中国”“美国”“德国”等;2)“不能表示所属学科专用概念的不足以反映实质内容的词”,如“方法”“问题”“试验”“研究”“分析”等;3)加了限定词的复杂短语或是未经公认的缩略词等,如“新型甘蓝型油菜品种”“AP”等,它们虽然能反映文章主题,但不利于计算机检索、识别。

3 10个无效关键词计量学分析

3.1 不同年份科技论文中10个无效关键词频率分布

结合已有研究及编校实践经验,本文针对第2类无效关键词展开研究。分别选取“应用”“对策”“问题”“管理”“设计”“措施”“影响”“现状”“研究”“发

展”这10个科技论文中出现频率较高的无效关键词,限定文献来源为科技论文,在中国知网进行检索。10个无效关键词在2015—2019年发表的科技论文中的频率(以检索记录表示)及其累计词频(10个无效关键词检索记录条数之和与当年发表的科技论文总篇数的比值)分布如表1所示。由表1可见,整体看,不同年度无效关键词频率分布相对稳定,按检索记录数排序,从高到低依次为“应用”“对策”“问题”“管理”“设计”“措施”“影响”“现状”“研究”“发展”,呈现出明显规律性。2017年,科技论文发表总量为2 017 727篇,这10个无效关键词检索记录之和达82 732条(未排除有2个或多个无效关键词同时出现在同一篇科技论文中的情况),每100篇文章里就有约4篇文章的关键词里包含这10个无效关键词之一,可见无效关键词的使用相当普遍。2015—2019年,“应用”这一关键词出现频率最高,且其检索记录在2017和2018年分别达20 267和20 203条,明显高于其他9个无效关键词的出现频率,说明科技论文作者选择“应用”作为关键词的比例较高。纵向来看,2015—2019年这10个无效关键词检索条数随年度的变化并无明显规律性,累计词频范围为3.7%~4.1%,其中2015和2019年的无效关键词累计词频均为3.7%左右(表1),说明科技论文作者并未有意地减少无效关键词的使用,无效关键词将会伴随着科技论文作者的写作习惯而长期稳定地存在。

表1 2015—2019年科技论文中10个无效关键词频率及累计词频分布

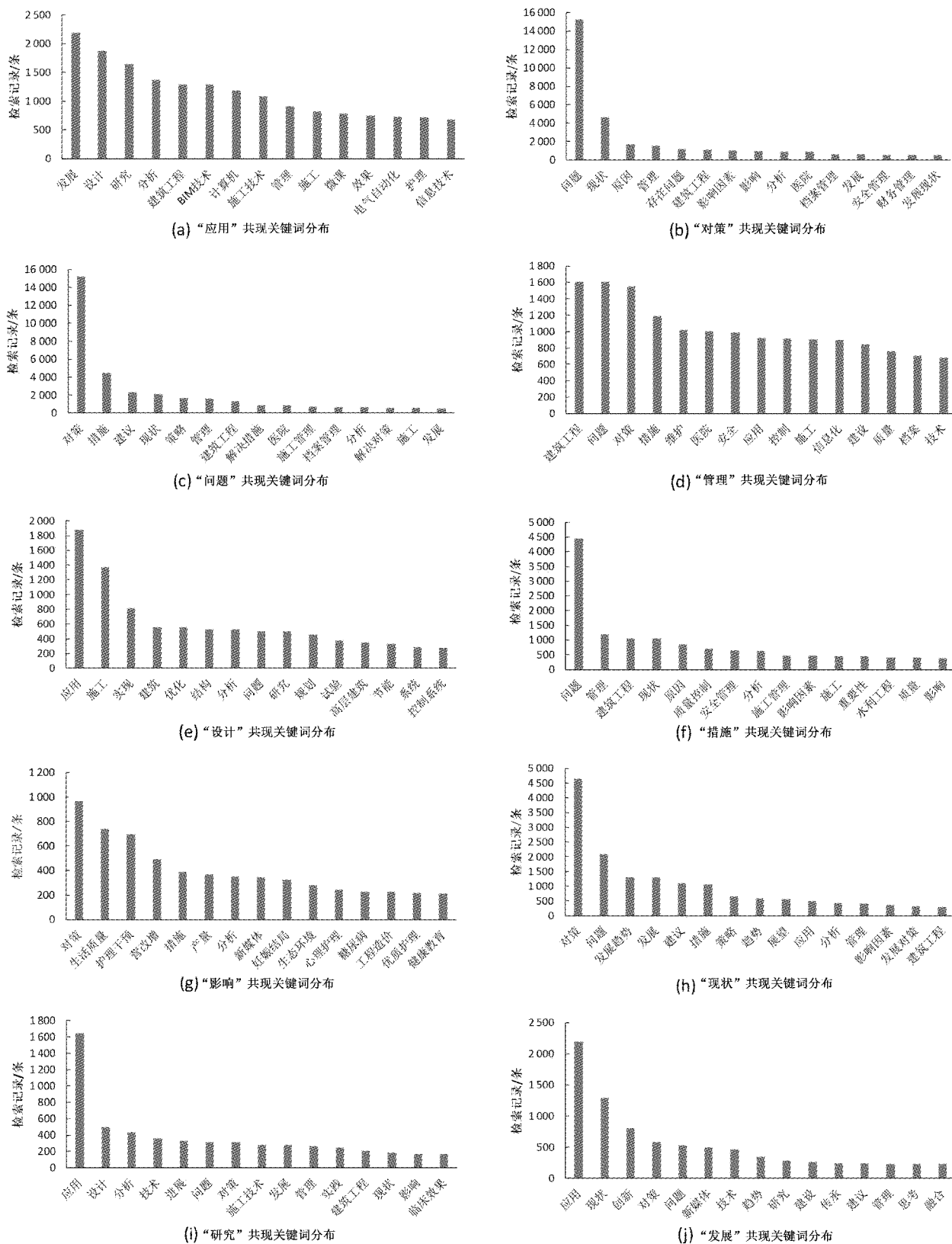
年份	总发文 篇数	无效关键词检索记录/条											累计 词频/%
		应用	对策	问题	管理	设计	措施	影响	现状	研究	发展	总计	
2015	2 056 986	16 364	10 715	8 368	7 984	7 454	6 453	4 679	5 029	4 670	3 717	75 433	3.7
2016	2 026 187	18 833	10 834	8 818	8 123	7 529	6 372	5 317	5 255	4 806	3 891	79 778	3.9
2017	2 017 727	20 267	10 827	9 447	8 122	7 786	6 954	5 493	5 067	4 746	4 023	82 732	4.1
2018	1 949 004	20 203	9 605	8 028	6 950	6 851	5 814	5 704	4 589	4 230	3 708	75 682	3.9
2019	1 840 114	18 644	8 560	7 216	6 161	5 870	5 022	5 126	4 041	3 843	3 187	67 670	3.7

3.2 10个无效关键词共词分析

共词分析法是指通过分析共同出现在同一篇文章中的关键词或主题词,从而鉴别某一学科知识结构和研究热点的方法^[7-8]。目前,基于关键词的共词分析法已被广泛应用于文献计量、信息系统、人工智能等领域^[9]。

本文以10个无效关键词为研究对象,分析其在

2015—2019年共现关键词(即2个关键词同时出现在同一篇文章中的现象^[10])分布情况,结果如图1所示。图1中,2015—2019年10个无效关键词“应用”“对策”“问题”“管理”“设计”“措施”“影响”“现状”“研究”“发展”的检索记录总数分别为94 100、50 287、41 644、37 074、34 989、30 565、26 275、23 905、22 225和18 448条。



数据来源于中国知网,本文仅展示频率排序第1~15的共现关键词。

图1 2015—2019年10个无效关键词共词分析

从图1可以看出,2015—2019年同一篇科技论文中同时出现2个无效关键词的现象相当普遍。例如,

以“问题”为检索关键词,与其共现频率最高的关键词为“对策”,共现频率达15 223(图1(c)),也就是说,

在一篇以“问题”为关键词的科技论文中,有大于1/3的概率会同时出现“对策”这一关键词。分析其余9个无效关键词的共词分布情况也可发现,与这些关键词共现频率较高的往往为无效关键词。例如,按照关键词共现频率排序,2015—2019年与关键词“应用”共现的无效关键词依次为“发展”“设计”“研究”,共现频率分别为2 188、1 878和1 642(图1(a));与关键词“现状”“研究”“发展”共现频率较高的几乎均为“对策”“问题”“应用”等无效关键词(图1(h)~(j)),可见无效关键词共现网络不仅不能反映学科知识结构和研究热点,而且造成了信息资源的浪费。值得一提的是,在本文选取的10个无效关键词排名前15的共现关键词中,有7个共现关键词中出现了“建筑”或“建筑工程”,这可能与该领域的科技论文发文总量较大有关:以“建筑”为主题词进行检索,2015—2019年累计发文306 196篇,约占科技论文总发文量9 890 018篇的3%。

3.3 不同期刊科技论文中10个无效关键词频率分布

以本文筛选出的10个无效关键词为检索条件,同

时限定文献来源为“学院学报”和“大学学报”,检索得到2015—2019年不同类型期刊发表的科技论文中的10个无效关键词频率分布及其累计词频分布如表2所示。由表2可见,2015—2019年,发表于“大学学报”的科技论文明显比“学院学报”的科技论文多,但“学院学报”科技论文中无效关键词的出现频率明显比“大学学报”科技论文的高。例如,2015年,“大学学报”共检索到67 598篇文献,其中以“应用”为关键词的文献为100条;“学院学报”共检索到42 620篇文献,其中以“应用”为关键词的文献为305条,二者差异显著。由表2还可看出,2015—2019年,“大学学报”科技论文中的无效关键词累计词频分别为0.73%、0.72%、0.69%、0.58%和0.54%，“学院学报”科技论文中的常见无效关键词累计词频分别为3.36%、3.34%、2.89%、2.65%和2.56%，二者均随着年度的变化而逐渐降低,呈现出明显的规律性。同时,“大学学报”科技论文无效关键词累计词频明显比同期“学院学报”科技论文的低,说明前者关键词标引质量比后者的高。

表2 2015—2019年不同类型期刊科技论文中的10个无效关键词频率及累计词频分布

年份	文献来源	总发文篇数	无效关键词检索记录/条											累计词频/%
			应用	对策	问题	管理	设计	措施	影响	现状	研究	发展	总计	
2015	大学学报	67 598	100	86	50	26	80	12	42	33	34	28	491	0.73
	学院学报	42 620	305	273	111	103	164	65	146	105	99	59	1 430	3.36
2016	大学学报	66 830	109	91	43	29	54	7	51	48	29	22	483	0.72
	学院学报	39 718	265	241	101	91	153	71	144	100	96	65	1 327	3.34
2017	大学学报	64 923	104	65	38	14	62	9	44	39	41	30	446	0.69
	学院学报	33 480	242	160	87	59	126	39	69	62	66	59	969	2.89
2018	大学学报	63 082	90	54	31	19	59	9	23	33	29	22	369	0.58
	学院学报	33 140	228	161	71	45	100	33	61	65	67	46	877	2.65
2019	大学学报	60 131	86	51	26	15	44	6	37	27	21	14	327	0.54
	学院学报	31 716	198	126	65	54	105	28	69	58	50	60	813	2.56

3.4 不同学历作者所著科技论文中10个无效关键词累计词频分布

以本文筛选出的10个无效关键词为检索条件,同时限定文献来源为“硕士论文”和“博士论文”,检索2015—2019年不同学历作者所著科技论文中无效关键词频率分布及其累计词频(即10个无效关键词检索条数之和与“博士论文”和“硕士论文”总篇数的比值)分布,如表3所示。

由表3可见,2015—2019年“硕士论文”总篇数显著高于“博士论文”总篇数,前者为后者的8.1~10.5倍;同时,“硕士论文”中10个无效关键词出现频率明显比“博士论文”的高,如2019年“硕士论文”中以“设

计”为关键词的检索记录为203条,而“博士论文”中的检索记录仅为1条。由表3还可见,2015—2019年“硕士论文”中的无效关键词累计词频明显降低,分别为10.43%、8.84%、8.32%、6.93%和5.32%,与2015年相比,2019年“硕士论文”无效关键词累计词频降幅达48.9%。2015—2019年“博士论文”中的无效关键词累计词频整体上也呈下降趋势,分别为2.39%、1.70%、1.71%、1.57%和1.40%,5年间“博士论文”无效关键词累计词频降幅达41.4%。同时,“博士论文”中无效关键词累计词频明显比同期“硕士论文”中的低,说明无效关键词的出现频率与论文作者学历有一定关系。

表3 2015—2019年不同学历作者所著科技论文中的10个无效关键词频率及累计词频分布

年份	文献来源	总发文篇数	无效关键词检索记录/条										累计词频/‰	
			应用	对策	问题	管理	设计	措施	影响	现状	研究	发展		总计
2015	博士论文	24 317	16	4	3	5	7	0	7	2	8	6	58	2.39
	硕士论文	213 902	301	371	149	167	555	43	188	130	219	107	2 230	10.43
2016	博士论文	25 816	14	3	2	3	4	0	6	0	8	4	44	1.70
	硕士论文	216 742	283	307	114	178	449	32	145	139	154	116	1 917	8.84
2017	博士论文	26 844	10	4	4	0	9	0	8	1	7	3	46	1.71
	硕士论文	218 093	280	312	115	143	450	34	127	116	147	90	1 814	8.32
2018	博士论文	24 160	12	5	1	2	3	0	4	2	6	3	38	1.57
	硕士论文	210 779	251	245	102	101	317	18	123	118	94	91	1 460	6.93
2019	博士论文	14 956	5	3	2	2	1	0	6	0	2	0	21	1.40
	硕士论文	156 545	163	144	30	53	203	14	77	62	46	41	833	5.32

4 10个无效关键词出现原因及应对措施

4.1 10个无效关键词出现的原因分析

综合分析结果可知,虽然在2015—2019年博士和硕士作者群体已逐渐减少无效关键词的使用,但是在2015和2019年发表的科技论文中,这10个无效关键词累计词频并无明显变化(约为3.7%),说明无效关键词存在此消彼长的现象。本研究认为无效关键词出现的主要原因包括以下几个。

一是多数科技论文作者未接受有效的指导,不知道如何撰写合适的关键词。有研究表明,多数高校学生承认自己未接受过系统的科研写作训练^[4,11-12]。屈李纯等^[4]针对328名硕士研究生展开调研,发现仅有17.39%的研究生熟悉关键词的选取。其实早在1987年,我国GB/T 7713—1987《科学技术报告、学位论文和学术论文的编写格式》明确规定:每篇论文应选取3~8个词作为关键词,如有可能,尽量用《汉语主题词表》等词表中提供的规范词语^[13]。2019年,国家新闻出版署发布了针对关键词的行业标准CY/T 173—2019《学术出版规范 关键词编写规则》,其中阐述了关键词的定义和基本要求。遗憾的是,高等院校中有关科技论文写作的课程和培训仍然较少,很多作者甚至从未听说过《汉语主题词表》,其他企事业单位中的作者接触到最新的标准、规范、写作培训和指导的机会更少,这就导致了低质量科技论文和无效关键词的稳定存在。例如,在未包括“质量”“分析”“技术”“原因”等无效关键词的情况下,2015—2019年,仅本文选取的10个无效关键词检索记录总数达381 295条;进一步分析这些无效关键词的累计词频可发现,其整体呈先增加后降低的趋势,但变化幅度不大,可见科技论文作者使用无效关键词的现象依然普遍。

二是部分作者未从思想上重视关键词对于论文检索的重要作用。有些中文科技论文作者只在乎论文是否能发表,是否能助其拿到学位或晋升职称,而对论文

的关键词标引质量以及论文发表后的学术影响并不关心,这就导致科技论文中存在大量无检索价值的关键词。进一步分析图1还可以发现,作者在选取关键词时,往往只是根据刊物要求,简单地将论文题目拆分为3~8个关键词,而未仔细鉴别这些关键词是否能够准确地反映该篇论文的特点,导致同一篇论文中无效关键词共现的情况也相当普遍。例如,2015—2019年篇名中同时含“应用”和“研究”的科技论文检索记录达199 841条,篇名中同时含“影响”和“研究”的科技论文检索记录达128 358条,这也反映了部分作者从论文标题中选取关键词的随意性。

三是论文审读编校群体(包括论文指导教师和同行评议人员)把关不严。目前,仍然有相当一部分的编校人员秉持着“改错不改好”的理念,在审读科技论文时仅仅关注变量符号和单位使用不规范、图表和公式编排不合理等,而针对论文关键词等能体现文章亮点的内容就抱着无所谓的态度,虽然也是“为他人做嫁衣裳”,但难免有不够走心之嫌。例如,2015—2019年来源于“学院学报”的科技论文总数为180 674篇,其中10个无效关键词的累计词频为14.80%;而来源于“大学学报”的科技论文总数为322 564篇,其中10个无效关键词的累计词频为3.26%(表2),前者累计词频约为后者的4.5倍,可见“学院学报”的编校群体对关键词的标引质量控制严格程度比“大学学报”编校群体的低。同时,高等院校教师忙于申请项目、开公司,疏于指导学生进行学术研究及写作的现象依然存在,论文作者因学术不端而被撤稿或是撤销学位的报道也屡见不鲜,可见高等院校中部分导师和同行评议人员对学生学术论文的主体内容审读尚不够仔细,遑论论文中的关键词了。例如,2015—2019年“博士论文”总数为116 093篇,其中10个无效关键词的累计词频为8.78‰;而“硕士论文”总数为1 016 061篇,其中10个无效关键词的累计词频为39.84‰,后者约为前者的4.5倍(表3),可见“硕士论文”的审读群体对

关键词的标引质量控制严格程度比“博士论文”审读群体的低。

4.2 无效关键词的应对措施

我们认为降低科技论文中的无效关键词出现频率的措施主要包括以下几个方面。

一是进一步加强对科技论文作者的指导与培训。这就要求高等院校、各科技期刊编辑部以及其他企事业单位形成合力,建立合理的培训机制,定期开设写作课程或开展写作培训,对作者进行系统而全面的科技论文写作训练,使其加深对论文结构、论文要素、论文写作标准及规范的认识,提高综合写作能力,减少论文无效关键词的选取。例如,吉林大学《中国兽医学报》编辑部针对科技论文写作者开展“移动课堂”网络教学,取得了较好的效果^[12]。

二是科技论文作者应从思想上重视关键词对于文献检索的重要作用,同时在操作上严格执行关键词的编写规则。作者应充分认识到关键词是科技论文检索的信息点,是表达论文主要学术观点的关键性因素,在论文写作中应避免使用无效关键词。CY/T 173—2019规定:关键词编写一般包括论文审读、主题分析、选词和编排;关键词应准确并充分揭示论文主题内容,重要的可检索内容不应遗漏^[2]。科技论文作者根据编写规则选取适宜的关键词后,还应以读者的身份进行反推:要检索这一研究领域的科技论文,是否会选择这样的关键词?合理、科学地标引关键词有助于提升文献利用率。

三是论文审读编校群体需对无效关键词进行严格把关。关键词能够反映期刊研究热点和发展方向,关系着期刊的学术质量^[14-15]。论文审读编校群体应重视关键词的标引工作^[16],选取合适的关键词有利于学术文献的有效传播和充分利用,从而有助于提高论文和期刊的学术影响力^[15]。例如,已成功入选“中国科技期刊卓越行动计划梯队期刊”的《中南大学学报(自然科学版)》严把期刊学术质量关,2015—2019年共发表论文2 385篇,其中10个无效关键词的出现频率为0;而《湖南城市学院学报(自然科学版)》同期共发表论文933篇,其中10个无效关键词的检索记录为68条,累计词频达7.3%。可见,论文审读编校群体在关键词的标引质量控制过程中发挥着重要作用。

5 结束语

在科学技术日新月异的今天,通过关键词检索相关研究领域的最新文献早已成为科研工作者最常用的检索方式之一。一方面,科技论文作者应主动学习科研写作标准和规范,参加科研写作课程学习或培训,根

据论文主题选取恰当的、指向性强的关键词;另一方面,作者和审读编校人员都应从思想上重视关键词的标引,进一步加强对关键词标引质量的控制,避免使用无效关键词,以期实现科技论文快速而广泛的传播,提高科技论文的检索利用率和学术影响力。

6 参考文献

- [1] 吕小红,杨开英,张蕾. 大数据时代精细加工提高科技论文显示度[J]. 编辑学报, 2018, 30(4): 373
- [2] 学术出版规范关键词编写规则:CY/T 173—2019[S]. 北京: 国家新闻出版署, 2019
- [3] 张媛,李春英. 无效关键词:现状、成因及对策:基于“中国知网”的高校社科学报论文分析[J]. 山西师大学报(社会科学版), 2015, 42(6): 156
- [4] 屈李纯,霍振响. 科技论文关键词“不关键”原因探析[J]. 编辑学报, 2019, 31(5): 516
- [5] 张银凤,高嵩,尹春霞,等. 学术论文中关键词的选取与常见问题分析[J]. 黄冈师范学院学报, 2019, 39(6): 223
- [6] 唐宏伟. 学术论文关键词标引的检索意义[J]. 青海大学学报(自然科学版), 2008, 26(1): 91
- [7] 张勤,马费成. 国外知识管理研究范式:以共词分析为方法[J]. 管理科学学报, 2007(6): 65
- [8] 储节旺,郭春侠. 共词分析法的基本原理及EXCEL实现[J]. 情报科学, 2011, 29(6): 931
- [9] 吉亚力,田文静,董颖. 基于关键词共现和社会网络分析法的我国智库热点主题研究[J]. 情报科学, 2015, 33(3): 108
- [10] CALLON M, COURTIAL J P, LAVILLE F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry[J]. Scientometrics, 1991, 22(1): 155
- [11] 张丽,黄琴,张旋,等. 高等医科院校开设科技论文写作课的实践探索[J]. 昆明医科大学学报, 2015, 36(4): 171
- [12] 刘珊珊,王浩然,孙晓芳,等. 高校学报“移动课堂”教学促进科技论文写作发展探索[J]. 编辑学报, 2020, 32(1): 101
- [13] 科学技术报告、学位论文和学术论文的编写格式:GB/T 7713—1987[S]. 北京: 国家标准局, 1987
- [14] 廖锡庆,李晓艺. 坚持文化导向 打造精品科技期刊:《编辑学报》2008—2018年载文关键词分析[J]. 科技传播, 2019, 11(24): 1
- [15] 李秀霞,程结晶,邵作运. 文献关键词分布特征与期刊学术质量的关系[J]. 中国科技期刊研究, 2019, 30(4): 426
- [16] 王昌度,熊云,徐金龙,等. 科技期刊论文关键词标引的问题与对策[J]. 编辑学报, 2003, 15(5): 349

(2020-03-19收稿;2020-05-30修回)