

科技论文作者选取关键词行为分析*

——以《计算机系统应用》为例

孙君艳 张重毅 方梅

中国科学院软件研究所期刊联合编辑部,100190,北京

摘要 为了分析科技论文作者选取关键词的行为和态度,研究科技论文作者对关键词检索的习惯,探讨优化关键词的方法,笔者设计了一套关键词选取方法调查问卷,涉及关键词检索习惯、选取方法、处理态度及期望等内容,并向《计算机系统应用》的作者发放200份调查问卷,收回有效问卷152份。结果显示,作者更倾向于采用主题或关键词检索模式进行论文检索,71.62%的作者重视关键词的选取,76.97%的作者仅从题目和摘要中选取关键词。73.65%的作者认为关键词是为了体现文章内容及技术要点,79.73%的作者研究过热点关键词,89.19%的作者认为有效关键词能提高文章显示度,93.10%的作者愿意编辑部推荐关键词。由于作者缺少有效的关键词选取方法,导致他们心有余而力不足。科技期刊编辑应在优化关键词方面进行深入研究并开展工作,与作者共同努力,使选取适当关键词更规范、有效。

关键词 科技期刊;行为分析;关键词选取;文献检索;影响力;研究热点

Analysis of keyword selection of authors from scientific papers: taking *Computer Systems and Applications* as an example//SUN Junyan, ZHANG Zhongyi, FANG Mei

Abstract This paper aims to analyze the behavior and attitude of authors from scientific papers towards keyword selection, investigate the habit of authors in keyword retrieval, and explore the method of keyword optimization. We design a questionnaire about the keyword selection, including keyword retrieval habit, selection method, processing attitude, expectation, etc. Two hundred questionnaires are sent to the authors of *Computer Systems and Applications* and 152 valid ones are retrieved. In comparison with other modes, authors are inclined to retrieve papers in the subject or keyword mode. 71.62% authors pay much attention to keyword selection. 76.97% authors select keywords only from titles and abstracts. 73.65% authors believe that keywords are set to reflect the article content and technical points. 79.73% authors have investigated hot keywords. 89.19% authors think that valid keywords can increase article display. 93.10% authors hope editorial boards to recommend keywords. Since authors are lack of effective method, they are powerless in keyword selection. Editors of scientific and technological journals should make in-depth studies in keyword optimization and make efforts together with authors to select more standard and valid keywords.

Keywords scientific and technological journals; behavior

analysis; keywords selection; literature retrieval; impact; research hotspots

Authors' address Editorial Office of Joint Journals, Institute of Software, Chinese Academy of Sciences, 100190, Beijing, China

DOI:10.16811/j.cnki.1001-4314.2022.05.010

随着科学技术的不断发展,国家对中文科技期刊的重视程度逐渐提高^[1],每年都有大量的科研论文发表在中文期刊上。中国知网(CNKI)数据显示,2019年发表的129万篇科学技术专业领域论文中发文量最高的学科是“自动化技术、计算机技术”(87 673篇,6.79%)^[2]。计算机类的论文如何在大量的文献中脱颖而出,被更多的学者发现,除了文章质量和研究方向以外,规范、准确的关键词标引也是需要关注的一个方面。《科学技术报告、学位论文和学术论文的编写格式》^[3](以下简称《编写格式》)指出:“关键词是为了文献标引工作,从报告、论文中选取出来用以表示全文主题内容信息款目的单词或术语。”关键词是文章内容的浓缩,反映了文章的核心创新点。借由关键词,读者可迅速了解文章的主要内容。有效关键词就是能够准确表述文章主要内容,有效展示科研工作关键点的词语。另外,从《编写格式》可知,关键词的设置是为了文献标引工作,有助于读者更精准、高效地检索论文。热点关键词则是根据文章内容选择领域研究热点,有助于提高文章被检索的概率,提升阅读兴趣的词语。因此,选取有效且契合研究趋势的热点关键词,在提高文章可读性、提升显示度方面的作用尤为重要。

然而,由于作者对关键词标引作用的了解相对较少,期刊论文中关键词的选用仍存在许多问题,针对此现象,许多学者进行了研究^[4-10]。例如,张红霞等^[5]从信息传播视角探讨了学术论文关键词标引的方法和策略,提出了关键词标引的步骤:确定主题概念、主题概念转换、关键词审核修改;屈李纯等^[6]分析了科技论文关键词不关键的原因,认为科技论文关键词“不关键”的原因有关键词标引缺少指导规范、作者对关键词标引不熟悉以及期刊编辑“忽视”关键词编校;伍锦花等^[7]分析了科技论文中10个不能反映论文主题的低效关键词,如“对策”“问题”“影响”,并给出了相应的原因和应对措施。

* 中国科学院科学出版基金(传播字[2020]5号)

中国科学院软件研究所期刊联合编辑部承担三刊一网工作,即《软件学报》《计算机系统应用》《International Journal of Software and Informatics》和中国计算机期刊网。笔者在《计算机系统应用》日常编辑审稿工作中发现,作者在选取关键词时存在重复选取、使用不规范、内容不全面、选取不准确等问题。科技期刊编辑作为论文的加工者,有责任帮助作者完善文章内容,不断提高出版质量,提升文章显示度。从长远来看,单从某篇文章出发来辅助作者修改关键词不具有可行性,应找到文章中存在的共性问题,并了解作者需求,从编辑部和作者两方面来考虑,既可以让作者更清楚地了解关键词标引的规范,又可简化编校过程。为全面了解作者在关键词选取方面的习惯和需求,笔者以《计算机系统应用》为研究样本,在实务工作中围绕关键词提炼并设计了一套调查问卷,问卷内容包括关键词检索习惯、选取方法、处理态度及期望等内容,以期了解作者需求,掌握科技论文作者在关键词处理方面的行为特征,为后续开展关键词推荐和分析工作提供参考。

1 研究方法

1.1 问卷设计

笔者采用问卷星制作本次调查问卷。问卷共有11题,包括排序题和单选题,从检索方法、关键词选取方法、对关键词的态度以及作者对编辑部推荐关键词的看法等角度设计问卷,问卷的题目类别、详细内容、题型和题目数量如表1所示。通过邮箱或官方微信公众号共发放200份问卷,截至2022年7月15日,共收回有效答卷152份,答卷有效率为76%。

表1 调查问卷题型设计

类别	内容	题型	题目数量
检索方法	检索网站	排序题	3
	检索模式		
关键词选取方法	中英文全称、缩写习惯	单选题	3
	关键词来源		
	关键词数量		
对关键词的态度	热点关键词研究情况	单选题	3
	关键词重视情况		
作者对编辑部推荐关键词的看法	作者对有效关键词的看法	单选题	3
	作者对关键词作用的想法		
作者对编辑部推荐关键词的看法	推荐关键词的意愿	单选题	2
	推荐关键词的数量		

1.2 分析工具

本文使用问卷星数据统计和分析功能进行数据分析。在作者检索方法行为分析时,采用选项平均综合得分来反映作者的排序情况,选项平均综合得分的计

算公式为

$$a = (\sum p \times w) / n,$$

式中: a 为选项平均综合得分; p 为选项被选择的频数; w 为选项的权值,由选项被排列的位置决定; n 为本题填写人次。例如,有3个选项参与排序,那么排在第1、2、3个位置的权值分别为3、2、1。如果一个题目共被填写10次,选项A被选中并排在第1个位置5次,第2个位置2次,第3个位置3次,那么选项A的平均综合得分 $= (5 \times 3 + 2 \times 2 + 3 \times 1) / 10 = 2.2$ 分。由此可知,选项平均综合得分的最大值为选项的个数。

2 结果分析与讨论

2.1 结果分析

2.1.1 作者的检索习惯 关键词是为了方便论文检索而设计的,因此调查作者的检索习惯有利于从根本上了解读者对关键词的看法。表2为根据作者的检索习惯统计得到的平均综合得分情况。可以看出,选择中文学术数据库(如中国知网和万方数据知识服务平台)和学术搜索引擎(如百度学术和谷歌学术)进行检索的平均综合得分最高,分别为4.82和4.49,排名第3、4、5的分别为国际学术数据库、国外学术出版机构和期刊官网。作者主要选择主题和关键词来检索论文,使用期刊名和作者进行检索的相对较少。在全称和缩写的使用上,使用中文全称进行检索的平均综合得分最高,为4.14,其次为英文全称,使用英文缩写、中文全称+英文缩写、英文全称+英文缩写的比例依次减少。

表2 作者检索习惯调研

题目类型	选项	平均综合得分	说明
检索网站	中文学术数据库	4.82	根据公式(1),选项数为6个,因此平均综合得分最大值为6
	学术搜索引擎	4.49	
	国际学术数据库	2.99	
	国外学术出版机构	2.05	
	期刊官网	1.70	
检索模式	其他	0.18	根据公式(1),选项数为5个,因此平均综合得分最大值为5
	主题	4.31	
	关键词	3.89	
	期刊名	1.47	
	作者	1.34	
中英文全称、缩写习惯	其他	0.04	根据公式(1),选项数为5个,因此平均综合得分最大值为5
	中文全称	4.14	
	英文全称	3.21	
	英文缩写	1.49	
	中文全称+英文缩写	1.45	
	英文全称+英文缩写	0.91	

综上所述,大部分作者选择中文学术数据库和学术搜索引擎进行文献检索,检索模式主要为主题和关

关键词,并且多采用中文全称或英文全称进行检索。

2.1.2 作者选取关键词的方法

表3为作者选取关键词的来源。76.97%的作者从题目和摘要中选取关键词,11.18%从相关文章中选取,9.21%通过查询相关学科术语词表进行选择,2.63%通过导师建议来选取关键词。在数量方面,选取3~4个和5~6个关键词的人数相近,分别为51.35%和48.65%。对于热点关键词的关注方面,有79.73%的作者表示考虑过研究热点关键词,但还有20.27%的作者未曾考虑过研究热点关键词。

表3 作者选取关键词的来源

来源	选择人数	占比/%
题目和摘要	117	76.97
相关文章	17	11.18
导师建议	4	2.63
相关学科术语词表	14	9.21

由此可知,题目和摘要是作者选取关键词的主要来源,选取3~4个和5~6个关键词的人数基本相当。另外,大部分作者在撰稿过程中考虑过研究热点关键词,但不能确认所选择术语是否为当前研究热点。

2.1.3 作者对关键词的态度 针对作者对关键词的重视情况,调研结果显示,71.62%的作者表示重视关键词的选取,会花费一定的精力去仔细考虑选取合适的关键词;28.38%的作者认为关键词只需要基本达到要求即可。至于有效关键词是否有提高文章显示度和影响力的作用,89.19%的作者认为是可以的,并且很需要关键词来判断研究兴趣的相关度,9.45%的作者表示没有考虑过这个问题,1.35%的作者不关注关键词。另外,73.65%的作者认为关键词最重要的作用是体现文章内容及技术关键点,12.16%的作者认为关键词是为了引导读者在阅读前了解文章内容,14.19%的作者认为关键词是为了方便检索。93.1%的作者表示愿意编辑部向其推荐关键词,说明作者渴望得到编辑部的支持。对于推荐的数量,以为1~2个为佳的比例为55.56%,认为3~4个比较合适的占25.93%,18.52%的作者对推荐关键词的数量没有要求。

由此可知,大部分作者重视关键词的选取,并且认为有效关键词能够提升文章的影响力。另外,他们认为关键词最大的作用是体现文章内容及技术关键点,引导读者在阅读前了解文章内容。90%以上的作者愿意编辑部推荐关键词,55.56%的作者认为推荐关键词的数量为1~2个较为合适。

2.2 讨论

2.2.1 作者青睐的检索网站较一致 根据调研结果

可知,中文学术数据库和学术搜索引擎的平均综合得分最高。另外,从后台详细填写数据也可以看出,把中文学术数据库和学术搜索引擎作为第一检索来源的人数分别为67人和62人,合计占比84.9%,说明大多数作者检索论文时优先考虑中文学术数据库和学术搜索引擎。这一方面是因为中文学术数据库和学术搜索引擎收录的论文数量多,文章类型全面,作者可下载到大部分自己感兴趣的论文;另一方面是因为《计算机系统应用》是中文科技期刊,作者倾向于选择中文学术数据库进行参照学习。

2.2.2 主题和关键词是最主要的检索选项

调研结果显示,作者主要通过主题和关键词进行检索,而通过期刊名和作者进行检索的比例较少,这说明科技论文作者对主题和关键词的关注黏性比对其他选项的黏性更大,他们更关注研究的主题和方向,而对某一期刊或作者关注度不高。这也侧面提示编辑部在编校和专题策划中要多关注文章主题和关键词的选择,从而提升文章的显示度。

2.2.3 中英文全称为作者的检索首选 在关键词的书写上,中英文关键词应尽量提供中英文全称,从而让更多的读者检索到该篇文章。例如,在CNKI数据库检索发现,以卷积神经网络(convolutional neural network, CNN)为研究主题的文章,用“卷积神经网络”为关键词检索,学术期刊和学位论文的文章数量为3.25万,而以“CNN”为关键词检索出来的数量为1.13万,前者检索文章数量是后者的3倍左右。由此可见,以中英文全称作为关键词能让文章更大程度被其他学者检索到。

2.2.4 从题目和摘要中选取关键词过于片面 结果显示,作者主要从题目和摘要中选取关键词,显然这一方法是不全面的,因为关键词是全文内容的浓缩,仅仅出自题目和摘要,可能会遗漏主题信息^[11]。另外,值得一提的是,作者对于学科术语词表的关注不够,只有8.11%的作者在撰稿时会参考学科术语词表,这表明作者对于关键词的规范性缺少关注。需要注意的是,关键词标引应严格遵守规范性、全面性、专指性和客观公正性的原则^[12-13],规范的关键词不仅有利于读者检索,也有利于编辑部进行文献主题信息挖掘^[14]。

2.2.5 热点关键词值得关注 通过统计关键词、主题词、篇名词等核心词汇在某一类学术文献中出现的频次,可判别该学术领域的研究热点、知识结构和发展趋势^[15-16]。虽然79.73%的作者都研究过热点关键词,但从实际效果来看,他们对热点关键词的了解不够,导致选取的关键词不够“热点”。编辑部作为文章的把关者,有能力也有方法总结热点关键词库供作者选取,

从而让他们的文章达到更好的宣传效果。

2.2.6 关键词的检索作用未引起作者的重视 调研结果显示,85.81%的作者认为关键词最重要的作用是体现文章内容及技术关键点,引导读者在阅读前了解文章内容,而这些内容在题目和摘要中已经能够详尽地体现,显然这不是关键词最重要的作用。如《编写格式》所述,关键词最重要的作用是文献标引,便于读者检索。从结果来看,只有14.19%的作者意识到关键词的检索作用。可见,关键词的设置目的和作者对关键词的认知之间存在偏差。另外,从学者的检索习惯来看,他们在检索文章时大多采用关键词进行检索,而在撰写文章时对关键词的检索作用反而重视不足,这也是编辑部需关注的一个矛盾点。

2.2.7 有效关键词有助于提高文章影响力 有效关键词,顾名思义,是对文章检索和阅读起有效作用的关键词。编辑部可在关键词的长度、书写规范以及全称、缩写的使用等方面辅助作者修改关键词,让关键词更有效。根据调研结果显示,仍有10.8%的作者不了解也不关注有效关键词的作用,这一点也是编辑部今后需关注的一个方面,即提升作者对有效关键词的重视,以共同提升文章质量及显示度。另外,值得注意的是,17.57%的作者认为有效关键词可以提升文章影响力,但却不重视关键词的选取,说明科技论文作者的思想认识与实际行动之间是有差距的。从前文分析可知,这是因为作者缺少合适的关键词选取方法,进而导致了他们心有余而力不足的状态。

2.2.8 编辑部应设计适当方法为作者推荐合适的关键词 调查结果显示,93.10%的作者表示愿意编辑部向其推荐关键词,表明了作者对编辑部的信任和期望。另外,编辑部根据长期的工作积累经验,可以设计一个以规范化关键词、推荐热点关键词为目的的方法。

3 结束语

通过调研期刊作者对关键词的方法和态度来了解关键词方面目前存在的问题。关键词作为论文检索的关键因素,其标引的规范性和全面性将直接影响文章的检索显示度。结果显示,科技论文作者大多从中文学术数据库通过主题或关键词进行论文检索;一般从题目和摘要中选择3~6个关键词,整体符合期刊关键词数量要求;大多数作者重视关键词的选取,也希望通过关键词来判断研究兴趣的相关度,但由于方法和经验有限,作者普遍希望编辑部能够推荐合适的关键词来提升文章的影响力。

在下一步的工作中,编辑部将进行2方面研究:一

是研究关键词对文章显示度的量化作用;二是利用智能化数据分析工具,建立学科领域热点关键词库,探索优化关键词选取的途径。

感谢期刊联合编辑部祁丽娟副编审和牛欣悦编辑在问卷设计和调研过程中给予的大力支持。

4 参考文献

- [1] 国家新闻出版署. 中共中央宣传部、教育部、科技部印发《关于推动学术期刊繁荣发展的意见》的通知[A/OL]. (2021-06-23)[2022-08-15]. <https://www.nppa.gov.cn/nppa/contents/312/76209.shtml/>
- [2] 中国科学技术协会. 中国科技期刊发展蓝皮书(2021)[EB/OL]. (2022-01-27)[2022-08-15]. <https://stm.castscs.org.cn/lpsxz/38714.jhtml>
- [3] 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:中国标准出版社,1990
- [4] 张建蓉,陈燕. 学术论文中关键词标引的常见问题剖析[J]. 编辑学报,2003,15(2):104
- [5] 张红霞,冀论文,贾丽红,等. 信息传播视角下学位论文关键词标引方法与策略[J]. 中国科技期刊研究,2018,29(12):1213
- [6] 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
- [7] 伍锦花,陈灿华. 科技论文10个无效关键词计量学分析[J]. 编辑学报,2020,32(4):403
- [8] 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
- [9] 吴军. 新媒体时代期刊关键词的应用探析[J]. 出版广角,2021(11):49
- [10] 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
- [11] 杨珍珍. 学术论文关键词的选取[J]. 兰州工业学院学报,2018,25(5):122
- [12] 蔡益群. 基于《中国分类主题词表》电子版的学术论文关键词标引[J]. 图书情报导刊,2018,3(6):1
- [13] 程智强,江洪涛. 科技论文关键词标引应琢磨的几个问题[J]. 编辑学报,2014,26(1):40
- [14] 边钊,唐娉,闫珺. 关键词规范化对文献主题信息挖掘的影响:以遥感领域为例[J]. 中国科技期刊研究,2021,32(12):1535
- [15] 张勤. 词频分析法在学科发展动态研究中的应用综述[J]. 图书情报知识,2011(2):95
- [16] 池营营,安珍,周小潭. 基于载文关键词统计的《编辑学报》研究热点分析[J]. 传播与版权,2019(1):87
(2022-08-24收稿;2022-09-22修回)