

低重复率稿件中的学术不端行为检测与防范

吴昔昔 贾建敏 吴健敏 王小同[†]

温州医科大学期刊社,325035,浙江温州

摘要 重复率低却存在抄袭的稿件,其原因可能为单一的检测系统存在漏洞,新发表的论文上传数据库存在时间差,作者对稿件的有意识“修饰”,以及网络时代造假手段多样化。为此,文中提出以下措施:使用多种数据库进行文字检测,从查重报告中发现问题,建立有效的审稿专家数据库、网站教育和警示,充分利用网络资源,加快稿件处理进程,增强编辑的责任意识等,以利于更好地防范学术不端行为。

关键词 期刊稿件;学术不端;查重;防范策略

How to detect and prevent academic misconduct in manuscripts of low detection rate // WU Xixi, JIA Jianmin, WU Jianmin, WANG Xiaotong

Abstract The paper analyzes manuscripts of the low detection rate with suspected plagiarism. There are several possible reasons: using a single detection system has its flaws; there is the time gap between newly published papers uploaded to the database; the author modified the manuscripts intentionally; fraudulent measures in the internet are diverse. It suggests that the manuscripts of low detection rate need to be paid more attention to: firstly, using a variety of databases for text detection; finding problems from the report; establishing an effective review expert database; enhancing web education; making full use of network resources; speeding up

学术价值,才能提高期刊的学术质量,提升期刊的学术影响力。在当前的科技期刊界,很多专家都身兼几种期刊的编委。丁佐奇等通过调查《中国天然药物》的编委,发现一半以上的专家同时担任超过5种期刊的编委,最多的一位担任了26种期刊的编委^[3]。在这种大环境中,如何吸引编委把更多的注意力和精力投入期刊建设,发挥更加重要的作用,是值得研究的课题。编辑部应该尊重、肯定编委,为编委搭建宣传平台,提供更多的优质服务,同时注重对最新科研知识的推送,激发编委更大的热情,使编委的作用最大化,实现期刊与编委的共同发展。

5 参考文献

- [1] 刘岗,魏海明,王婉,等. 编委在科技期刊发展中的作用探讨[J]. 中国科技期刊研究,2015,26(3):239
- [2] 王金莲. 高校学报编委会作用的最大化[J]. 编辑学报,2012,24(6):573
- [3] 丁佐奇,郑晓南,吴晓明. 编委贡献源于编辑部做好编委

the working process; last but not least, strengthening the responsibility of editors. Above suggestions would be effective measures to be taken to prevent academic misconducts.

Keywords manuscript of journal; academic misconduct; detection; prevention measure

Authors' address Editorial Department of Wenzhou Medical University, 325035, Wenzhou, Zhejiang, China

DOI:10.16811/j.cnki.1001-4314.2016.03.021

学术不端行为在论文发表中普遍存在,由于具有较强的隐蔽性,防范和查处这些学术不端论文具有一定的难度^[1]。反学术不端行为是编辑人员审稿过程中的一项重要内容^[2]。在实际工作中,学术不端文献检测系统功能强大,通过对稿件进行文字查重,提供相关参考文献进行对比分析,帮助编辑对稿件做出正确判断,极大地提高了编辑工作效率,有效地打击了学术不端行为;然而,学术不端文献检测系统也存在一些漏洞,检测结果只能作为参考^[3],我们需要客观、合理地对待检测结果。通过学术不端文献检测系统查重后,显示与其他已经发表的论文的重复部分少的稿件,即

服务工作:以《中国天然药物》为例[J]. 编辑学报,2011,23(4):339

- [4] 朱晓文,宋冠群. 编委会对提高期刊质量所起的作用[J]. 编辑学报,2013,25(6):564
- [5] 汤敏华. 充分发挥编委作用 办好中医药期刊[J]. 今传媒,2012(12):120
- [6] 姚巍,朱金才. 关于审稿质量和最小发表周期的关系[G]//学报编辑论丛. 上海:上海大学出版社,2005:184
- [7] 郝丽芳,陈宏宇,武文. 科技期刊通过自我宣传提高学术影响力的途径:以《生物工程学报》的实践为例[J]. 编辑学报,2015,27(3):285
- [8] 王亚秋,陈峰,王家暖,等. 强化编委职能 实现科技期刊可持续发展[J]. 编辑学报,2011,23(3):244
- [9] 王丽萍,郑志惠. 探讨如何更好地发挥编委会作用[J]. 新闻研究导刊,2015(2):66
- [10] 吴爱华,王晴,杜冰,等. 科技期刊应重视对审稿专家的知识回报[J]. 编辑学报,2013,25(2):164
- [11] 胡英奎,游滨,王秀玲,等. 科技期刊编辑跟踪学术前沿的途径[J]. 编辑学报,2010,22(4):307

(2015-10-20 收稿;2016-01-18 修回)

为低重复率稿件。部分低重复率稿件也有可能存在学术不端行为,应该引起高度重视。

本文就编辑在查重过程中遇到的一些重复率低却存在抄袭嫌疑的稿件进行原因分析,并提出相应建议,以便于编辑人员更好地把关。

1 稿件查重结果

作者投稿后,对属于期刊刊载范围内的稿件,编辑会先进行文字重复率检测。检测结果分为以下2种。

一种是重复率高的稿件。一般认为文字重复率高于30%的稿件存在抄袭嫌疑,通常会直接退稿;但是,有些稿件还是要编辑根据具体情况做出客观、合理的判断。如稿件系来自研究生论文(尚未发表过)中的部分内容,或是作者毕业后同一课题的系列研究的后续论文,可以考虑通过查重检测。此外,医学论文中涉及的诊断标准、疗效标准、方法研究、疾病概念和药物成分,在表述的过程难免也会有重复,一般可以不认为是抄袭。

另一种是重复率低的稿件。这类稿件通常会有2种情况:一是稿件的确有创新性,观点正确,且语言表达规范,是具有理论价值或实用意义的好文章;二是稿件的文字重复率虽然低,但是却存在学术不端行为。

2 稿件重复率低的原因分析

2.1 单一的检测系统存在漏洞 有研究者^[4]认为学术不端检测系统是存在缺陷的,这主要受制于计算机开发软件,其开发者的目的主要是甄别出机械的抄袭和拼凑行为,为各类论文提供简单的检索。一些变相抄袭的论文较容易被系统误判,得到较低的重复率。中国知网(CNKI)、万方数据是国内最具有影响力的2大论文数据库,它们各自都收录了大量的文献。由于受到独家授权的影响,数据来源和版权保护方面竞争激烈,两者的收录覆盖面可能不一样^[5],也有所侧重,使得检测结果颇有差异^[6]。大部分论文都被CNKI和万方收录;但是有的论文只被CNKI收录而未被万方收录,或者相反;也有可能是某一篇论文都被CNKI和万方收录,但时间有先后:所以2大数据库之间的内容有重合,又难免各有遗漏。因此,当同一篇稿件在同一时间、同一地点用AMLC和万方论文相似性检测系统进行检测时,其检测结果可能有所不同,甚至个别的差异还比较大。例如《血清游离核小体……研究》一文,用AMLC检测后查重率为6.1%,万方检测结果为66.7%;而另外一篇《FOCUS-PDCA程序……的应用》,AMLC检测的重复率为40.2%,而万方则为5.9%。

此外,不同文本格式的稿件,其查重结果差异也比

较大,公式、图表和外文重复率问题也未能解决^[7]。

2.2 新发表论文上传数据库存在时间差 新出版的论文与其被数据库收录存在时滞。如新收稿件若与新发表且尚未被数据库收录的论文内容相似,则在文献检测时就较难发现问题,也就有可能检测不出稿件中的抄袭部分。由于不同刊物间的信息难以实现共享,也使部分作者钻了空子。如笔者遇到过一篇稿件《二维斑点……的改变》,2014年1月2日初次查重的结果重复率为20.1%,而在2014年10月27日再次检测时,其重复率变为34.0%。

2.3 作者对稿件有意识“修饰” 由于作者的刻意为之,一些稿件中的内容的确存在学术不端的问题;但文献检测系统却很难检测出重合部分,使得编辑人员在初审过程中容易被这类稿件所蒙骗。

1)全文雷同。抄袭者掌握了检测系统的查重原理,对所抄袭的论文在文字方面进行了一定的处理。通过对抄袭部分的文字进行“加工”,如增删文句、拆分或者合并段落,改变语句或表达方式,把他人的论文“改装”为自己的“论文”,从而顺利逃过检测系统的查重。

笔者收到过一篇稿件《树脂修复……的研究》,AMLC查重结果为4.1%,万方为0;但是编辑还是发现它与论文《后牙病理性磨损树脂修复的临床效果探讨》雷同。这2篇文章在摘要、引言、各级层次标题以及图标方面几乎一样。如《后牙病理性磨损树脂修复的临床效果探讨》在引言部分的一句话是“牙齿磨损是常见的一种非龋性损害,已成为影响口腔健康的又一重要问题。成年人牙齿磨损发生率接近100.0%”,而《树脂修复……的研究》则将此句改为“作为一种常见的非龋性损害,牙齿磨损已成为危害口腔健康的一个重要因素,多见于成年人,发生率几乎达到100%”。全文其他部分几乎是意思式抄袭。

2)核心内容相似。剽窃他人的观点,套用他人的设计思路来发表论文的情况也屡见不鲜。有些作者出于同事间的情谊,让他人将自己已被录用的稿件稍作修改后投稿。这类稿件虽然采用了作者自己的文字表述,数据上也有所不同,但全文的宗旨、核心内容和观点并未改变。笔者在初审一篇稿件时发现,其题材和内容与本刊即将发表的另外一篇稿件十分相似,只是在例数上有差别。经查实,这2篇稿件的作者属于同一个单位的同一个科室。

3)部分结果套用。由于检测系统对一些图表几乎无法检测^[4],一些作者将已经发表的论文中的图表重新包装后投向其他杂志社^[8],如将图表改成文字叙述,或者将文字叙述改为图表,这种情况下单用检测系

统就很难查出其抄袭部分。作者将同一篇稿件投了多家杂志或者将几篇内容相似的稿件投给不同的杂志。更有甚者,还有除第一作者外的其他作者将同一篇稿件(尚未发表)的内容修改后再投稿的情况。如本刊一名责任编辑将一篇待发排的稿件进行二次查重后,发现稿件中的第二作者已经在几个月前以第一作者的名义在其他杂志上发表一篇论文,该文中有2/3的数据和图片来自那篇论文。经调查,该文的第一和第二作者是夫妻档。

2.4 网络时代造假手段多样化 2005年,美国麻省理工学院研究人员发现了SCIgen这个“神器”软件,它能结合字符串产生假文章^[9]。其初衷是为了证明假论文更容易被会议接受,却不想扰乱了本就已经混乱的学术论文市场。不少学术不端的作者钻了空子,一些通过计算机合成的假论文不断涌现出来。网络科学技术的飞速发展以及造假工具的频繁出现,使得一稿多投和抄袭剽窃行为变得更加容易和隐蔽。

3 应对策略

3.1 使用多家数据库进行文字重复率检测 目前国内单一数据库的期刊收录不全,覆盖率不高,可能较难百分之百地检测出稿件中的重复部分。一篇稿件的多种查重结果有利于编辑人员对稿件做出合理的判断。如今有越来越多的数字出版商开发了新型的数字资源平台,不仅可以为编辑的学习提供资源空间,而且免费提供了文献检测。期刊编辑部不妨加盟它们,这些检测平台相对于传统数据库(CNKI、万方、维普等),可以实现专题相关文献和学术观点一站式呈现,有利于加强对学术不断的防范。此外,可以根据不同稿件内容选择合适的数据库进行查重。如中华医学会系列杂志仅在万方数据库发行,与CNKI数据库相比,万方数据库中的医学类期刊占有明显的优势^[10],而在人文科学方面却稍显不足。例如《温州医科大学学报》,是一种综合性医学学术期刊,接收稿件的内容主要以基础医学、临床医学以及预防医学为主,少数为教学研究类论文。大部分医学类的稿件多采用万方数据库来检测,若遇到人文研究类稿件,则首选CNKI数据库进行检测。如一篇关于高教类的论文《基于混合……模式实践》,CNKI的检测结果为25.7%,而万方的检测结果为3.7%。

3.2 从检测报告中发现问题的 对部分查重率低的稿件需要特别注意,尤其应该谨慎对待其检测报告。我们可以从这些检测报告中发现端倪。

1) 重视“稿件已经提前检测”。有些检测报告上会显示“稿件已经提前检测”,这种情况的出现有以下

可能:作者“一稿多投”,其他编辑部已经对此稿件进行了检测;作者“改装”他人论文后自行多次检测修改以便于逃过编辑部的查重。如上文提到的《树脂修复……的研究》一文,就因其检测报告中出现了“稿件已经提前检测”而引起了编辑的怀疑。遇到这类稿件,我们需浏览相关领域的论文进行查实。

2) 关注参考文献的重复比。稿件查重报告中的重复比(红色区域)大部分出现在参考文献中,编辑可搜索这些红色文献的来源,将其与所检稿件进行仔细对比,判断所检稿件与所查文献的相似度。笔者初审稿件《连续性非卧床……的影响分析》时,发现该文中的参考文献有90.0%来自《维持性血液透析与腹膜透析患者血压变异性比较》一文。编辑部将这2篇文章从内容到参考文献都进行认真核对,最终认为该文没有创新点而退稿。

3) 观察所引文献的题名。检测报告会将所检稿件中涉及的文献(包括已引用或未引用)一一列出。若编辑察觉到这些文献中有题名与所检稿件相似的,可以将其下载后细读。如本刊编辑在查看稿件《复方丹参注射液对……的影响》时,无意中看到与其题名类似的文章《复方丹参注射液对羊水过少胎儿氧化应激及妊娠结果的影响》。阅读这篇文章后,发现除了部分数据不同,所检稿件与该篇文章有相同的层次标题和图题以及内容相差不大的讨论。

3.3 建立有效的审稿专家数据库 编辑人员的专业知识范围十分有限,加之稿件被人为修饰,有部分隐蔽抄袭内容是检测系统和编辑人员不易察觉的。审稿专家是遏制学术不端行为的最有效的力量^[11]。由于待审稿件多、时间和精力耗费大,或部分稿件不对口,使得有些审稿专家缺乏严谨的审稿态度,对稿件审读草草了事。而对稿件研究领域熟悉的专家有利于提出合理、中肯的意见,也可以更好地对稿件进行把关。扩充专家库来建立有效审稿机制非常必要,足够的专家资源既可以减轻部分专家的审稿负担,更可以确保同行评议的详细和正确。

笔者认为,可以通过以下方法增加审稿专家人数:在数据库或高教图书馆中输入关键词,搜索相关研究领域的论文,查看里面的通信作者简介;直接在所检稿件的参考文献中寻找合适的审稿专家;在作者投稿时,建议他们提供“推荐审稿专家”和“回避审稿专家”以及联系方式。

3.4 网站教育和警示 汪勤俭等^[12]提出期刊编辑部应该主动防范学术不端,加强对作者的宣传教育和监督。如今大多数期刊编辑部都拥有自己的投稿网站,可以将本刊的稿约、论文写作规范、出版道德法规等内

容放到网站首页的下载中心里,供作者在投稿前阅读和学习。为了警示那些有学术不端念头的作者,编辑部可以将学术不端的处罚条例(内容包含警告、撤稿、列入黑名单、告知作者单位等)放入公告中。同时,可以将一些学术不端行为的案例也放入网站中或者借助微信平台发给作者,时刻起到警戒的作用。

3.5 充分利用网络资源 编辑可以在 CNKI、万方和维普数据库中通过搜索稿件中的关键词,查看是否有相似论文出现,再对相似论文的摘要部分进行对比;也可以通过高校图书馆数据库中一键搜索或者跨库搜索同类论文,浏览主要篇目。一般具有抄袭嫌疑的稿件,尽管在文字方面经过刻意修改^[13],但是摘要作为文章的核心部分,其内容较难改变。如《子宫……临床分析》这篇稿件,则是通过在数据库中输入“子宫切口妊娠”等这几个关键词,搜索到《对 35 例子宫切口妊娠临床治疗分析》,这 2 篇文章在内容和设计上十分雷同。此外,编辑人员也可借助别的检测工具,如百度、谷歌、搜狗等搜索引擎进行排查。

3.6 加快稿件处理进程 大多数期刊编辑部的来稿量较大,期刊所能发表的版面十分有限,难以满足全部作者发表论文的需求。一篇稿件从投稿到出版本就需要一定的周期,编辑部的稿件增多,更是延长了稿件的出版周期。作者在焦急等待的过程中,容易将稿件投向多个杂志以提高录用率。为了减免这种情况的发生,编辑部可以在自身方面做一些事情。比如编辑可以在作者投稿后要求他们提供单位盖章、作者(稿件中的每一位作者)亲自签名的相关证明(投稿介绍信、版权协议或其他有效协议)。虽然这样的要求使作者投稿的流程变得烦琐,但是正是因为程序复杂了,一些作者也不愿意轻易改投他刊。送专家审稿要及时,并在稿件审回后,尽早通知作者进行下一步的工作,尽可能地缩短稿件的发表时滞。

3.7 增强编辑的责任意识 由于检测系统的结果与论文真实相似度会存在一定差异,且虽然审稿专家是论文科学价值的主要评审者;但是一些隐藏的学术不端行为容易被忽视,这就需要编辑有足够的耐心去找出这些问题。在稿件“三审”中,初审环节是防范学术不端行为的必要“关口”,必须增强编辑的责任意识。稿件经过论文数据库查重后,建议对有嫌疑的稿件要进行深入了解,除了采用检测系统进行文字对比,还可以对低重复率的稿件通过各种检索手段进行进一步的排查。这些流程大大增加了编辑的工作量,也是对编辑的责任心的一种考验。同时,责任编辑应该对相关

的研究现状和学术动态十分熟悉,注意专业知识的学习,树立终身学习的观念,重视审稿经验的积累,从而及时发现学术不端行为,使有真正学术价值的论文得到发表。

4 结束语

学术不端文献检测在编辑初审工作中作为一种辅助工具,有效的杜绝论文抄袭现象,是编辑人员不可缺少的“好帮手”。但是,编辑人员不可过度的依赖文献检测的查重结果,认清低查重率的稿件也可能存在抄袭嫌疑。坚守编辑的职责,注意增加自我的知识积累,浏览相关学科的学术文章,结合自身的学识和工作经历,利用现有的学术资源和网络技术手段,提高鉴别稿件能力,严格把关,尽可能的挖掘和制止隐藏的学术论文中的不端行为。

5 参考文献

- [1] 张和,鲁翠涛,张海燕. 学术不端文献检测系统在《肝胆胰外科杂志》中的应用[J]. 肝胆胰外科杂志, 2014, 26(3):262
- [2] 李新根,徐用吉. 学术期刊编辑如何防范学术不端行为[J]. 科技与出版, 2010(8):64
- [3] 谭华,崔洁. 学术不端文献检测系统的使用建议[J]. 编辑学报, 2010, 22(2):153
- [4] 朱银环. 刻意规避学术不端软件检测论文的再审查[J]. 编辑学报, 2015, 27(3):249
- [5] 张浩,高国龙,钱俊龙. 国内外学术不端文献检测系统平台的比较研究[J]. 中国科技期刊研究, 2011, 22(4):514
- [6] 倪爱香. 学术期刊编辑对学术不端行为的防范策略[J]. 中国科技期刊研究, 2013, 24(3):586
- [7] 何文辉,林海霞,张梅,等. 从责任编辑视角谈科技期刊学术不端检测系统的使用[J]. 编辑学报, 2013, 25(3):276
- [8] 王福军,谭秀荣,冷怀明. 科技期刊中常见的学术不端现象分析与思考[J]. 编辑学报, 2014, 26(5):452
- [9] 徐海丽. 学术不端行为及其预防措施[J]. 中国科技期刊研究所, 2015, 26(6):545
- [10] 曾华明. 国内三大中文期刊全文数据库述评[J]. 农业网络信息, 2014(2):544
- [11] 朱银环. 期刊主题人员防范深度学术不端行为的职责分析[J]. 中国科技期刊研究所, 2014, 25(11):1373
- [12] 汪勤俭,郭建秀,栾嘉,等. 84 篇因学术不端退稿稿件的追踪分析和思考[J]. 编辑学报, 2012, 24(2):131
- [13] 王新英,赵艳静,赵杨. 彻查科技论文学术不端的编辑策略[J]. 编辑学报, 2011, 23(3):519

(2015-12-30 收稿;2016-02-22 修回)