

论学术不端的深度防范

肖骏 谢晓红 王淑华

中国地质大学《地球科学》编辑部,430074,武汉

摘要 定量测试内容篡改和字序变更对学术不端检测系统(AMLC)查重的影响,发现字序更改和有无“触媒”都能极大地影响 AMLC 的检测结果。认为期刊编辑部应加强出版队伍建设,发挥编辑主观能动性和主人翁意识,合理使用 AMLC 及其检测结果,以达到深度防范学术不端,提高期刊质量的目的。

关键词 学术不端;学术不端检测系统(AMLC);编辑队伍;期刊质量

A discussion on deep prevention against academic misconduct after "DONG Peng plagiarism incident" // XIAO Jun, XIE Xiaohong, WANG Shuhua

Abstract Quantitative tests were taken to figure out how tampering content and changing word order influence the AMLC test results. In fact, the changes of word order and the existence of "catalyst" can greatly affect the AMLC test results. It is believed that sci-tech journal editorial offices should strengthen the construction of the publishing team, make the editors have initiative and sense of ownership, and use the AMLC and its testing results accurately to deepen the prevention against academic misconduct and improve the quality of journals.

Keywords academic misconduct; academic misconduct literature check (AMLC); editorial team; quality of journals

Authors' address Editorial office of Journal of Earth Science, China University of Geosciences, 430074, Wuhan, China

DOI:10.16811/j.cnki.1001-4314.2017.04.018

2016年9月,成都商报披露了“论文大神”董鹏大量抄袭剽窃论文事件。这个董鹏,自2011年起,5年发表了约800篇文章,涵盖物流、经济、美学、心理学、电影等多个领域。调查发现,董鹏伪造身份,虚假署名,虚构、篡改国家级科研项目,一稿多投在上百种学术期刊上。被“打假”后,他承认了抄袭事实并为自己的行为道歉^[1]。董鹏如何成功规避上百种学术期刊的学术不端检测,值得出版从业者深思。

目前,绝大多数期刊编辑部利用学术不端检测系统(AMLC)防范抄袭、剽窃、伪造、篡改等学术不端行为。AMLC利用自适应多阶指纹特征检测技术检测文章内容重复度,它的推广对一些学术不端行为有很强的震慑作用,极大地提高了编辑部排查学术不端行为的效率,其检测结果已成为期刊编辑部取舍稿件的重要依据。大量文献定性总结了使用AMLC的体会:确实能检测内容重复^[2-8]。那么,“董鹏事件”究竟是学术造假者越来越精明^[9],还是学术不端检测形同虚

设?调研发现,网络上有详细的逃避AMLC检测的攻略^[10],更有付费资源提供查重结果帮助作者精确修改规避期刊AMLC检测。另有传闻,简单变更字序也能干扰检测系统的查重结果。

笔者选择CNKI数据库学术论文中一段原文进行了系列测试,一来检测AMLC文字篡改后的抗干扰能力,二来建立对AMLC检测结果的定量认识,帮助期刊编辑部全面了解AMLC,有助于深度防范学术不端,提高期刊质量。

1 测试

截取文献[11]第2节第4点最后一段话进行了内容篡改(内容1)和不同密度的字序变更(内容2和内容3)。

原文 这种模式有其存在的现实可能性,经营得法则可做到社会效益和经济效益兼收;但困难也是显而易见的。比如:高校早已清理整顿校办企业,再申请成立企业性质的期刊出版公司,学校会有很多顾虑;一校一个出版单位,只出版1种或几种期刊,难以形成经营规模;再说,这种小规模出版单位不易获得批准;如果这种出版企业设立过多,政府主管部门将来也不好管理。

内容1 这种模式是可能实现其存在的,而且一旦经营得当则可做到经济效益和社会效益兼收;但显而易见,困难也是有的。首先,清理整顿过校办企业的高校再申请成立具有企业性质的期刊出版公司会顾虑很多;其次,一般情况下一所学校可能只有1种或几种期刊,难以形成有经营规模的出版企业;再说,政府主管部门考虑到将来可能会不好管理而不会轻易批准设立过多这种小规模出版企业。

内容2 这种模式有其存在的实现可能性,经营得法则可做到社会效益和经济效益兼收;但困难也是显而易见的。比如:高校已早清理整顿校办企业,再申请成立企业性质的期刊出版公司,学校会有很多顾虑;一校一个出版单位,只出版1种或几种期刊,难以形成经营规模;再说,这种小规模出版单位不易获得批准;如果这种出版企业设立过多,政府主管部门将来也不好管理。

内容3 这种模式有其存在的现实可能性,经营

得法则可做到社会效益和经济效益兼收;但困难也是显而易见的。比如:高校早已清理整顿校办企业,再申请成立企业性质的期刊出版公司,学校会有很多顾虑;一校一个出版单位,只出版1种或几种刊期,难以形成经营规模;再说,这种小规模出版单位不易获得批准;如果这种出版企业设立过多,政府主管部门将来也不好管理。

2 结果与讨论

原文共164字符(含标点符号),内容1保留原文120字符(含标点符号),内容2和3皆为完整原文。内容2变换字序30处,而内容3变换字序15处,仅更改了“模式”“出版”“企业”等少数词的字序。将截取的原文和修改后的3段文本放在一个文档中进行AMLC检测,检测出重复文字数分别为64、98、161,3段文字总重复度达到64.6%(表1);但将修改后的3段文字放在一个文档中则检测结果重复度为0%。

表1 含修改后文本及原文的查重结果(2016年10月25日)

查重项目	字符数	重复字数	重复度/%
内容1	172	64	37.2
内容2	164	98	59.8
内容3	164	161	98.2
合计	500	323	64.6

2.1 “触媒”效应 从测试结果很容易看出:不含原文的3段文字和含原文的4段文字AMLC检测结果截然不同。不含原文时,内容篡改和简单变更字序后的内容都能干扰AMLC的查重效果。而加入原文后,结构和词序都发生变化的句子仍能被标记为抄袭,改变字序则需要达到一定的密度才能避开查重检测。原文像“触媒”一样能引导AMLC顺势做出更正确的检测。更多检测表明,插入原论文中的任选段落时,AMLC都能识别3段文字的重复字段,几乎能“意会”内容1中某些句子。笔者还尝试用其他期刊文献原文作为“触媒”,但没有获得预期的结果,AMLC对篡改后的3段文字检测报告仍然为0%。由此看来,AMLC研发团队需设计能激发查重触发点的算法,使之发挥更高效的作用。

2.2 “字序”效应 目前,AMLC的指纹粒度不能设为单个汉字,不能识别字序错乱的重复内容;因此,简单更改内容字序通过检测的概率很大。很多期刊编辑部为提高稿件处理效率,会先查重后审稿,稿件重复度不低于设定值,就直接退回作者修改或拒稿,避免浪费编辑和审稿专家的时间。这种稿件处理流程容易让更改抄袭内容字序的文章通过学术不端检测系统查重而进入审稿流程。由于人脑的自动排序功能,会将某些错乱的字理解为“正确”^[12],审稿时常会忽略汉字和英

文字母顺序的变化,即便发现也可能认为是作者笔误写错。如果这种文章被录用发表,势必造成学术不端,影响期刊声誉,破坏正常出版秩序,甚至可能将期刊卷入司法程序。

3 对策

“字序”效应表明作者简单篡改抄袭内容都会对AMLC检测结果造成明显干扰,而“触媒”效应则显示如果能够准确触发AMLC查重活性,就能得出比较精确的检测结论;但在AMLC研发团队攻克“触媒”问题之前,期刊编辑部仍有许多工作要做。

3.1 合理使用AMLC及其检测结果 AMLC识别文字重复的功能是强大的,期刊编辑部应合理使用AMLC及其检测结果。合理使用AMLC包括同时使用多家数据库提供的文献检测系统和在稿件处理过程的不同阶段使用文献检测系统。期刊常用的有初审检测、刊出前检测、刊后检测等方式,初审检测是大部分期刊的首选^[13]。为使作者不能利用更改汉字顺序的方式避开AMLC查重,期刊编辑部应先通读来稿再做查重处理。初审通读时可借助拼写和语法检查软件,采用机器校对和人工校对相结合的方法。对于有太多拼写和语法错误的来稿,宜退回作者修改后再投。如果全文只是某部分错误较多,需要逐字阅读修改后再进行学术不端检测。应特别注意引言、讨论和结论等内容错误率高可能意味着存在学术不端行为。刊出前终稿的再次查重也能有效阻止学术不端的发生^[13]。可以排除审稿周期差异造成的重复发表,也能排除作者根据审稿意见修改和增补文章内容时剽窃、抄袭他人的学术成果。

AMLC的检测结果可以作为引导编辑排查抄袭、剽窃、篡改的线索。编辑部在使用AMLC检测结果时,不应仅局限于正文报告。文献^[14]对于使用AMLC检测报告进行了详细论述,提出重视“稿件已经提前检测”“参考文献重复比”“观察索引文献的题名”等,帮助编辑顺藤摸瓜发现学术不端。此外,编辑应分析每一处重复字段,将来稿和重复文字来源文章进行比对来甄别作者是否篡改了抄袭内容,干扰了检测结果。

3.2 加强期刊人力资源建设 防范学术不端行为需要大量精力和时间进行排查,而期刊竞争激烈,为提高效率,编辑部不断压缩稿件的出版周期,造成现今编辑部对文献检测系统过度倚重。量测试结果清楚显示,文献检测结果只能作为参考,不能作为决定性指标。此外,期刊还应以人为本,注重期刊人力资源的建设,维护科学诚信,维护出版秩序。

1) 激发编辑主人翁意识和责任心,调动他们的主

观能动性。不应以为有学术不端检索系统就可以高枕无忧,而应主动收集学术不端行为的信息,经常归纳总结学术不端稿件的特征,作为期刊编辑识别学术不端行为培训的资料,提高编辑对来稿内容的敏感性。积极改进有漏洞的出版流程,整合优化各类资源,深度防范学术不端。对于查重率低的稿件,也应从投稿者角度出发,借助来稿的研究方向、关键词、参考文献等信息,利用网络资源寻找学术不端的蛛丝马迹^[14-15],至少阻止“张卫”与“郭佳楠”这样匪夷所思的抄袭事件发生^[16]。同时,还应加强编辑科研积累,组织编辑跟进阅读科研论文,参加学术会议,了解学科最新研究进展。这样一支自省、创新、前进的编辑队伍在防范署名不当、同行评议不当等其他学术不端行为时也有极大的主动权。

2)从期刊审稿系统大数据中提取每位专家的审稿风格,利用 CNKI、Web of Science、ORCID、Research Gate、LinkedIn 等网络平台高效搜集专家信息,不断优化期刊编委、作者、审稿专家资源,建立一个知识结构全面且活跃的审稿专家数据库,用多样的审稿专家弥补编辑人员专业知识不足的缺憾,助力编辑部甄别学术不端行为。在数据库中将所有专家按研究方向和研究地区(如果有)分组,来稿通过编辑部初审和查重后,精准定向投送给特定的编委和专家组粗看一遍,尽力排除抄袭剽窃的可能后再选择审稿专家审稿。

3.3 调动全民防范学术不端 大量学术不端行为都由读者、同领域研究人员、被抄袭作者偶然发现或举报得以曝光^[1,16]。鉴于期刊编辑部工作繁杂,难以在时间和人力上全面保障学术不端防范,可在期刊主页、微信公众平台以及印刷本留白处发布“悬赏令”,调动关注期刊的所有人来帮助防范学术不端行为。对发现期刊录用和刊发文章有抄袭、剽窃、篡改等学术不端行为的人,经期刊编辑部鉴定认定事实成立给予相应的奖励。一经确认,对于实施学术不端行为的作者进行严肃处理,手段包括:1)将学术不端行为事实公布在期刊新闻主页、微信公众平台上;2)撤销学术不端行为责任人名下在本期刊上所发全部文章并列入黑名单;3)将事实致函其他刊发其文章的期刊;4)联系工作单位人事处并协商作者可能的经济赔偿以补偿期刊编辑部的各项损失。此类措施至今没有任何期刊编辑部实施,有很多的不可控因素,但值得尝试。所有措施和方法都由一点点的创新开始慢慢不断完善,有措施和方法总比束手无策。

4 结束语

深度防范学术不端,期刊编辑部不能全盘依赖学

术不端检索系统,要重视编辑和审稿专家的重要性,增强编辑部人工干预的力度;但根除学术不端行为还需努力加强作者学术道德建设,加重对学术不端行为的处罚力度,制订期刊出版道德规范,改进唯文章论英雄的评价机制,等等。挖去学术不端生长的土壤,避免“董鹏事件”再次发生。

5 参考文献

- [1] 张柄尧. 起底“论文大神”董鹏:文章是抄的,身份是假的[EB/OL]. [2016-11-24]. http://news.china.com.cn/2016-09/08/content_39256465.htm
- [2] 杨晨晨. 运用学术不端文献检测系统检测医学论文存在的问题及对策[J]. 编辑学报,2014,26(1):42
- [3] 柯文辉,林海清,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
- [4] 靖华. 科技期刊学术不端文献检测系统使用中的体会与思考[J]. 科技与出版,2011(7):60
- [5] 朱银周. 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
- [6] 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
- [7] 颜峻,侯风华,黄莉,等. 防范学术不端 净化高校学风:用“学术不端文献检测系统”的体会[J]. 编辑学报,2010,22(增刊2):8
- [8] 赵茜. 使用学术不端文献检测系统如何把握好合理使用的“度”[J]. 中国出版,2010(10):59
- [9] 廖雪丽,刘冰. 学术不端测谎仪科学性遭质疑[EB/OL]. [2016-11-24]. http://zqb.cyol.com/content/2009-04/22/content_2634086.htm
- [10] 怎么逃避知网学术不端检测系统[EB/OL]. [2016-11-24]. http://zhidao.baidu.com/link?url=1YXf_XJMosOdySuT2pcs0_rPIrdAaNmGhXkCozpOpN3p6xa_NyUAR1s_FxBlx24VDVsT-polBG0KK_scpqApGW6_Iomd-kRrwYlv_ihq0bKwNa
- [11] 张铁明,颜帅,赵大良,等. 高校科技期刊的出版模式初探[J]. 编辑学报,2010,22(4):328
- [12] 汉字顺序错乱并不影响阅读:网传“错句”引争议[EB/OL]. [2016-11-24]. <http://cq.qq.com/a/20130507/000116.htm>
- [13] 王音,田喆. 从编辑部公告看使用学术不端检测系统中存在的问题[J]. 编辑学报,2011,23(5):42
- [14] 吴昔昔,贾建敏,吴健敏,等. 低重复率稿件中的学术不端行为检测与防范[J]. 编辑学报,2016,28(3):266
- [15] 陆宜新. 学术不端检测中的隐性重复现象分析[J]. 中国科技期刊研究,2015,26(6):553
- [16] 一场匪夷所思、令人震惊的抄袭大案![EB/OL]. [2016-11-24]. http://mp.weixin.qq.com/s/OtvFFacDgq_OgVSty2CmoQg (2016-11-29 收稿;2017-01-19 修回)