

正确处理文字复制比为 0 和 >40% 的论文

赵 秋 民

《郑州大学学报》编辑部, 450001, 郑州

摘 要 中国知网研制的“学术不端文献检测系统”(AMLC)在编辑初审稿件时其检测结果会出现文字复制比为 0 和 >40% 这 2 种特殊情况。文章分析导致这 2 种检测结果的原因是: AMLC 检测不出来的论文文字隐性重复以及论文文字复制比 >40% 的论文中的格式性语言的重复、实验方法或操作步骤的重复, 疾病的病理生理机制、治疗方案、结果判定(或疗效)标准的重复, 以及统计学处理描述的重复, 并给出比较具体的解决方案。建议编辑人员在处理上述情况时不完全依赖检测结果, 要根据不同情况区别对待, 利用专业知识区分有无抄袭以及抄袭与不当引用, 正确使用 AMLC 提供的结果。

关键词 科技期刊; 学术论文; 学术不端检测; 文字复制比

Properly treat manuscripts with a copy ratio of zero or greater than 40 percent//ZHAO Qiumin

Abstract CNKI's academic misconduct literature detection system (AMLC) plays an important role in the first check of manuscripts for editors. Two detection results are very common, that is, the copy ratio is 0 and copy ratio is greater than 40 percent. The author analyzes the causes of the two kinds of results. AMLC cannot detect the invisible repeat in the manuscripts, and those repeats in format language, experimental methods or procedures, pathophysiology mechanisms of diseases, treatment methods, result judgment (or effect) standard and statistical analysis description with a copy ratio greater than 40 percent. Also, it cannot offer specific solutions. The author suggests that editorial staff not depend on the test results solely when handling the above circumstances, but should treat them according to different conditions, make use of professional knowledge to distinguish copy, plagiarism and inappropriate use, and adopt the AMLC results properly.

Keywords sci-tech journal; academic paper; academic misconduct detection; copy ratio

Author's address Editorial Board of Journal of Zhengzhou University, 450001, Zhengzhou, China

目前使用最为广泛的由中国知网研制的“学术不端文献检测系统”(AMLC)能够在 2~5 s 内完成一篇 5 000 字的文献比对并生成检测报告, 而且检测内容齐全, 详细列出文献与对比数据库重合文字的来源、发表日期与刊物、重合字数、文字重合度、引证关系、诊断类型等在内的检测结果, 大大节省了编辑检索与阅读文献的时间^[1], 该软件对期刊编辑初审论文、把握论文的学术水平及创新程度等提供了很大的帮助。在使用 AMLC 的过程中, 大多数编辑同人对学术不端的表现

形式都非常清楚, 但在实际操作时, 如何判断是否存在学术不端行为仍缺乏统一的尺度。近年来, 许多编辑部都认识到不能简单地以文字复制比的高低为标准来判断是否存在抄袭等学术不端, 而要结合重复的具体内容进行综合判断^[2-3]。笔者发现, 文字复制比为 0 和 >40% 这 2 种比较特殊的情况, 在此与编辑同人共同探讨这 2 种情况的正确处理办法。

1 论文文字复制比为 0 的情况

在实际工作中笔者发现, 有些论文文字的重重复 AMLC 是检测不出来的。这些文字复制比 AMLC 检测结果为 0 的文章, 其作者常将别人的论文改头换面, 或将别人的观点作一些语序调整、支离分解等, 或将要拷贝的文字用 google、百度等翻译工具译成英文, 然后自己再通过软件译成中文, 论文的意思大概相同; 但不同软件的翻译各不相同, 与被抄袭的论文拉开了相似度, 致使 AMLC 无法识别, 即文字复制比为 0, 检测通过。

还有一些作者通过网络搜索, 在投稿之前使用 AMLC 或其他学术不端检测软件, 将检测过的有学术不端的文字重新加工(如颠倒语序、改变句式、替换词语、增添一些可有可无的字词), 更有甚者, 直接在提示存在学术不端行为的文字中间添加空格键等等, 以期达到减少文字复制比的目的。

经过他们如此的“精心”处理, 即使文章主体结构照搬原文, 检测结果也可能为 0; 但其造假的本质依然存在, 只是变得更为隐蔽和不易被发现^[4]。

笔者用了本编辑部 2011-09-29—2013-09-18 的 1 440 篇文稿, 其中有 309 篇(21.5%) 的文字复制比为 0, 低于杨柱星等^[5]报道的 32.3%。根据本刊相关专业编委的审稿意见以及编委会定稿的结论, 309 篇文稿中通过终审并最后录用的不足 20 篇。笔者分析了这 200 多篇未被录用的文字复制比为 0 的文稿, 发现它们存在以下共同点。

1) 论文所涉及的专业包含的期刊未在 CNKI 数据库中收录或收录不完整^[6-7]。

2) 论文中存在大量的用符号编辑器编写的公式。

3) 论文的结果内容较多, 并且大多采用图片或表格的形式。

4) 其他。本编辑部有一篇文字复制比为 0 的文

稿,经过初审和1位同行专家的复审,认为该论文有新意;在送终审期间,另一位同行专家的审稿意见返回,详细地列出该文哪一部分是引用国外某篇论文的哪一部分,并给出外文文献的打印稿,还在相应部分做出标志;最后,通过与该文作者沟通,他承认是将几篇外文文献翻译拼凑而成,该文被退稿。靖华先生^[8]在工作中也遇到过类似情况。

2 论文文字复制比超过40%的情况

尤伟杰等^[3]认为下列情况不应判定为抄袭或重复发表:1)与已发表文献的重复比大于40%,但来自本人作为第一作者在学术会议交流过的文稿或学位论文;2)重复比介于20%~30%,但来自本人作为第一作者在期刊上已发表的论文,且论文的核心内容(临床资料、结果和讨论中的主要观点)不同。

董建军^[9]和王新英等^[10]也持相同的观点。他们在使用AMLC过程中发现,有许多与来稿相重复的文章属于作者之前撰写的学位论文,其内容并未正式发表,或是作者以前发表的一些论文,此来稿属于作者的后续研究成果。此外,虽然来稿有些内容与已发表的论文有重复,但这些内容都是引言、材料与方法或讨论部分的内容。

杨柱星^[5]从医学论文的特殊性方面阐述了上述几种特殊情况不宜判定为抄袭。

笔者曾遇到类似的一件事情:一篇博士论文经AMLC检测,文字复制比达到30%,研究生院将该结果反馈给研究生和导师,要求整改,否则将推迟答辩。详看复制比结果,主要集中在论文的“材料与方法”部分,且所重复的内容是与该导师指导的已经毕业的一位研究生的论文一致。该导师回复研究生院:“本人从事的课题是一项连续性工作,采用的实验仪器、实验方法大部分相同,其实验条件、操作步骤基本相同,可以重复,无须作文字修改。本人认为该学生已达到博士研究生毕业水平,准予参加毕业答辩。”最终研究生院接受了导师的意见,该研究生顺利毕业。

2.1 格式性语言^[5]的重复 临床研究类论文中一般都会有何时、何地、研究对象的类型、如何选取研究对象、研究对象如何分组、该研究的伦理学问题的处理,等等。

如一篇已在某刊发表的论文中有如下的文字:

×年×月×日至×年×月×日,选取×医院×科的住院或门诊的×患者×例,均经×检查确诊,按照×的原则分为×组,各组患者对该研究知情并同意,签署知情同意书。该研究得到医院(或科研单位)伦理委员会的批准并备案。各组患者的年龄、性别、疾病种

类、病程等经统计学处理,具有可比性。

按此格式撰写的论文,虽然是从另一角度进行了研究,并得出了具有创新意义的结果;但其格式性内容的文字复制比肯定超标。其实,该结果是AMLC得出的错误的结论。

2.2 实验方法或操作步骤的重复 医学领域的基础研究类论文中经常会用到比较成熟的实验方法和常用的仪器、试剂等,如免疫组织化学染色法、免疫细胞化学染色法、蛋白印迹实验、原位杂交实验、RT-PCR、荧光定量PCR等。这些实验方法均有成熟的操作流程,使用者有可能会根据自己的实验条件作出小的调整,但大部分均为规范的流程;所以在论文中不可避免地会出现较多的重复。

2.3 疾病的病理生理机制、治疗方案、结果判定(或疗效)标准的重复^[1,11] 绝大多数医学论文中或多或少都会涉及疾病的病理生理机制、药物成分及作用机制、疗效判断标准等方面的内容。一些疾病的疗效判断标准是专业领域内的金标准,不可替代且内容也较多,为体现科学的严肃性,作者在写作时必须原文引用。一些药物的主要成分、作用机制等已是公知公认的,而某些论文在进行分析讨论时又必须依据这些详细、具体的理论来阐明问题。由于这些内容具有专业性与权威性,因此大部分作者都会原文引用。

2.4 统计学处理描述的重复 医学类期刊大多对论文的统计学处理都有明确的要求,即在论文中必须注明使用的统计分析软件名称、版本,以及采用的统计学方法及检验水准。统计学方法大都具有共性、普适性和权威性。所有的统计学软件和方法贯穿于整个科研过程,并采用与之相适应的设计;所以,不管研究内容如何,只要科研设计中的统计学方法相同,在撰写论文时正确、合理的统计学用语一定会相同或相似。这样,AMLC检测结果有大量重复就不足为奇了。

3 AMLC检测结果中2种特例的处理方法

3.1 文字复制比为0论文的处理

1)编辑在初审阶段遇到这一情况时,可做出标记,然后送同行专家着重从学术创新或实用性方面进行专业审稿,最终根据专家意见对稿件进行合理处理。

2)采用其他学术不端检测软件重新检测。目前,由于中国知网和万方数据库争相与各期刊编辑部签订独家合作的协议,如果某编辑部签订这一协议之后,就势必造成另一数据库对该刊原始数据的缺失,也就造成了现今国内没有哪一家数据库的数据是完整的。基于此,各编辑部可以采用不同的学术不端检测软件(如PaperPass论文通行证、武汉大学ROST反剽窃系

统、万方论文相似性检测系统、维普通达论文引用检测系统)或利用“百度”“搜狗”等搜索引擎重新对论文的题名、关键词等进行检测,以判断检测结果的准确性。

3)对于有英文摘要的论文,笔者建议可以使用国外科技文献数据库的学术不端文献检测系统平台,如全球最权威的英文检测系统——Turnitin, CrossRef 的 CrossCheck, Blackboard 的 Safeassign, 爱思维尔的 PERK,以及马里兰大学的 The Plagiarism Checker、Plagiarism Checker 公司的检测平台、Article Checker 公司的检测平台、Plagiarism Search 公司的检测平台、Plagiarism Detect 公司的检测平台和 The Plagiarism 公司的检测平台等^[12]对英文题名、关键词和摘要进行检测。

3.2 文字复制比>40%论文的处理

1)格式性语言的重复。对于这类超标问题,在与作者沟通时将部分检测结果反馈给他们,并建议他们重新组织语句,以更加合理的、与以往报道不同的词语和语序来正确表达论文的内容。

2)实验方法或操作步骤的重复。一般情况下,经典的实验操作方法和步骤都不一定完全适合所有的实验,作者在实验的操作中会在此基础上作适当的改进,而在撰写论文时常会将该部分内容予以忽略,从而丢失了一些创新点;所以,应当建议作者在充分引用参考文献的前提下,将自己改进的内容一并写入论文中,以提升论文的创新性。

3)疾病的病理生理机制、治疗方案、结果判定(或疗效)标准的重复。笔者在处理文字复制比>40%的论文时发现 AMLC 在设计或使用上存在一定的缺陷。比如,论文中涉及有关疾病的病理生理机制的阐述时作者引用了文献,但 AMLC 给出的结果显示该段文字的文字复制比为50%左右。其原因是,作者在文中给出的引文能正确反映该内容,属于文献的合理使用;但由于文献发表时间较 AMLC 给出结果中的文献早,而且该文献在中国知网中未检索到。经与该文作者沟通后,将文献换为 AMLC 给出的文献后,顺利通过 AMLC。

4)统计学处理描述的重复。笔者所在编辑部对此采用的方法是:在完整描述统计分析软件的名称、版本的前提下,针对文中每一观察指标均给出所用的具体方法,并将二者有机地结合起来予以表述。如“采用 SPSS 10.0 进行数据的处理和分析,应用单因素方差分析比较对照组、改水组、高氟组3组间水氟、尿氟及血清雄激素结合蛋白、睾酮的差异,两两比较应用

LSD-*t* 检验;应用多重线性回归方法分析血清雄激素结合蛋白、睾酮与年龄及尿氟水平的关系。检验水准 $\alpha = 0.05$ 。”通过将常用的方法描述变成结合论文具体内容的具有鲜明特性的语句,不但避免了学术不端检测的高重复率,也使论文带上了自己的特殊符号。

4 结束语

综上所述,对于论文中出现的2种特例,编辑在初审下结论时一定要慎之又慎,应仔细阅读全文、对比论文中出现重复的文字段落,掌握重复的数量、重复内容的形式以及作者重复的目的,合理地处理发现的问题,利用专业知识区分有无抄袭以及抄袭与不当引用,正确使用 AMLC 提供的结果,利用自己锐利的眼光给论文作出合理、公正的评判。

5 参考文献

- [1] 江霞,方玉桂,陈伶俐,等. 医学科技期刊应用《科技期刊学术不端文献检测系统》判断抄袭时应慎重对待的几种情况[J]. 中国科技期刊研究,2010,21(4):471-472
- [2] 汪勤俭,郭建秀,栾嘉,等. 84篇因学术不端退稿稿件追踪分析与思考[J]. 编辑学报,2012,24(2):131-133
- [3] 尤伟杰,李小萍. 学术不端论文的一般规律及应对策略[J]. 编辑学报,2009,21(2):151-152
- [4] 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):153-155
- [5] 杨柱星,林荣慧,胡榜利,等. 医学论学术不端的特点及判定[J]. 中国科技期刊研究,2011,22(5):697-701
- [6] 胡朝明. 学术不端文献检测系统的使用与建议[J]. 图书馆工作与研究,2012(4):33-35,51
- [7] 胡政平. 学术不端文献检测与期刊质量的提升[J]. 出版发行研究,2012(11):87-89
- [8] 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):60-61
- [9] 董建军. 《眼科新进展》应用“学术不端文献检索系统”评价抄袭现象情况分析和体会[J]. 中国科技期刊研究,2011,22(3):388-390
- [10] 王新英,赵艳静,赵阳. 彻查科技论学术不端的编辑策略[J]. 编辑学报,2011,23(3):231-232
- [11] 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月上):59-61
- [12] 张旻浩,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):514-521

(2013-11-28 收稿;2014-01-03 修回)