

学术不端文献检测系统功能拓展^{*}

——基于编校流程的遴选审稿专家和监测参考文献

赵 蔚

《宁波大学学报》编辑部, 315211, 浙江宁波

摘 要 学术不端文献检测系统较多用于初筛环节检测文章的文字复制比,其基本原理是对文字进行比对和统计,基于比对,认为该检测系统也可以应用在审稿专家遴选和参考文献引用不实查证等方面。提出以文章重合部分检索潜在专家并建立可以共享的专家库,对不实引用参考文献实行标绿提示和统计标红参考文献字数的功能。

关键词 学术不端文献检测系统;评审专家;参考文献;不实引用

对于使用者来说,学术不端文献检测系统被普遍认可的功能是在初筛的环节检测文章的“文字复制比”,也就是标明文章的抄袭情况。在长期

的实践过程中,笔者发现学术不端文献检测系统的“比对”功能还可以拓展到编校流程的其他环节中,如系统在遴选审稿专家、监测参考文献是否如实引用等方面可以发挥作用,但要求使用者具备一定的技巧,并建议系统设计者对功能进行适度改进,以解决编辑流程中的新问题。

1 遴选审稿专家

1.1 稿件的新特征

随着科学技术的进步和研究方法的更新,研究者的科学研究视野早已经不再囿于单独的学科,综合化、交叉化、前沿化是目前科学研究的总体特征。^[1]以笔者所在的教育类期刊来说,最近几年的稿件越来越呈现如下特色:①稿件内容的多学科交叉现象非常普遍,经常有作者借用其他学科的理论分析本学科的问题;②作者的研究领域处于几个学科交叉的边缘地带,并且各学科以新的理论研究体系的方式组合在一起;③作者在某

^{*} 中国高校科技期刊研究会“2012年科技研究计划”:多学科视角下的学术不端文献检测系统及功能拓展研究(编号:GBJXC1231);浙江省高校学报研究会2012年重点课题:对学术不端与学术不端识别的多学科解读与干预(编号:ZGXB201205)。

度来审视,倡导邹韬奋本人“真诚为人民服务”的宗旨和信仰。作为一个行业的精神标杆,韬奋奖要给从业者传递一种人文情怀和价值取向,忠实地展现新闻专业主义的理念,不管市场化、世俗化和功利化等现象如何泛滥,都要保持超越的姿态,挣脱官本位的牢笼,重建一方明净的精神家园。唯其如此,韬奋新闻奖才能形成更深层次的道德魅力,获得更广泛的社会认可。

参考文献

[1] 雷群明.编辑记者要读点韬奋[J].编辑学刊,2008

- (5).
- [2] 顾子筠,张刚.曾立潮头沪学府:上海圣约翰大学[J].科学中国人,2007(4).
- [3] 马征,陈璐明.略论邹韬奋的新闻出版思想[J].新闻世界,2009(3).
- [4] 王瑞.从《店务通讯》看邹韬奋读者意识的演进[J].北京印刷学院学报,2008(5).
- [5] 王凤青.邹韬奋为何愤然辞去国民参政员[J].文史春秋,2009(7).
- [6] 徐新平.邹韬奋大众新闻思想述论[J].湖南大众传媒职业技术学院学报,2004(4). ■

领域的研究方向越来越前沿、尖端,也是目前学术论文的一大特色,每位研究者都希望自己研究的专题是最新的,以保证成果的独特性和创新性。

这些特点在学术上体现了百家争鸣、勇于创新的态势,但也直接导致研究领域越来越广、研究议题越来越分散,对编辑来说,这类稿件的审稿专家越来越难选。因为专家有一定的地域和专业约束,是相对较固定的。因此,前两种情况经常出现难于准确辨别某篇文章该送哪个学科更合适的情况,通常要送2~3个专家才能综合意见,辨别稿件的质量。后面一种情况要么专家对最新研究状况不是很清楚,不能很准确地判断稿件的创新程度,要么专家就要花大力气去了解这个专题的最新进展动态。目前,就作者所在学报来说,有将近1/4的稿件呈现这样的新特征,遴选最合适的专家成了越来越难以完成的任务。

1.2 精确确定审稿专家的方法

应用学术不端文献检测系统强大的文献比对库资源可以在更广泛的范围内遴选审稿专家,但不是所有的稿件都能应用该功能,要看稿件的文字复制比的情况。

1.2.1 待测稿件的文字复制比为0

在稿件初筛的过程中,要用到学术不端文献检测系统检测抄袭情况,如果稿件最终检测结果“文字复制比”为0,也就是没有任何抄袭,这是比较好的稿件(但也可能是更恶劣的意义改写抄袭稿件),这种文字复制比为0的稿件不能用于选择审稿专家。

1.2.2 待测稿件的文字复制比为1%~30%

如果该篇稿件的某段话被标红了,这说明可能存在抄袭的情况,如果文章的抄袭率在1%~30%之间,稿件有一定的学术价值,并且属于较难确定专家的稿件,就可以根据系统标红的部位,查看其性质。

(1) 如果被标红部分位于文章的前半部分,可以打开系统链接的被抄袭文章,查看其内容是不是相似的研究领域,因为同样的资料使用者研究的并不一定是相关内容课题,作者可能利用这个资料进行别的领域研究。如果是相似的研究领域,就可以按照这个线索,找到潜在的审稿专家。

(2) 标红的部位位于结论的部分,找到所需专家的可能性就更高了。如果链接的文章是所需的主题,就可以通过查看作者简介,进一步确定作者是否满足本刊审稿专家的条件。正规的期刊都有比较完整的介绍模式,如有完整的单位、职称、主要研究方向介绍,大部分都有电子邮件,这给我们寻找专家带来了很有利的条件,经联络确认后,就可以邀请其加入本刊的专家库请其审稿。

1.2.3 待测稿件的文字复制比为31%~100%

如果作者的抄袭率比较高,且没有多少学术价值的,可以直接退稿,但是笔者曾指出:系统收录不全、作者误引“抄袭文献”、引用了公有领域思想、文献综述类型所引起的误判^[2]导致文章超标现象要与作者核实,然后再作适度排除,如果达到了本刊通过审核的规定,可以遵照上面的方法查询审稿专家。

1.3 CNKI的审稿功能开发建议

在系统现有条件下,笔者认为CNKI的管理机构可以尝试在这个方面进行一些适度开发,完善和拓展系统的自身功能,这样可以更紧密地联系编辑的需求。基于查重链接的原理,笔者认为:

(1) CNKI可以将遴选审稿专家作为一项系统新增功能再度开发,引导期刊编辑更加注重作者的主要研究方向和联络方式。以文章重合部分作为检索项,将持有相同观点的文章链接并列列出来。系统还可以尝试更进一步,以主要研究方向和职称等作为筛选项目,帮助编辑选择所需专家。一旦编辑锁定专家,仿照设置自建比对库的原理,先开设总的专家库,然后容许编辑自建个性化专家库。

(2) 专门开设专家库,容许编辑上传经专家许可的信息,供全国的编辑筛选,这样可备筛选的专家范围大大拓展,能够在较高层次上实现精确的专家共享机制。

2 监测参考文献引用

GB/T 7714—2005《文后参考文献著录规则》将文后参考文献定义为“为撰写或编辑论文著作而引用的有关文献信息资源”^[3]。相信提起参考文献的校对难度,编辑们都是心有戚戚焉,下面从

编辑对参考文献的把关说起。

2.1 编辑对参考文献的把关

正确标引参考文献包括以下各个层面。

2.1.1 参考文献引用的合理性

《著作权法》对参考文献的合理引用有明确规定:①引用目的仅限于介绍、评论某一作品或者说明某一问题;②所引用部分不能构成引用人作品的主要部分或者实质部分;③不得损害被引用作品著作权人的利益;④应当指明作者姓名、作品名称。^[4]

针对上面的指导原则,笔者曾经撰文指出,少于10%被测文章的非“主要部分或者实质部分”的引用是合理的(综述类文章除外,这类文章最常见的是引用大量的参考文献),作者在文中正确标注即可。要注意的是,如果某作者大量引用某一篇或少数几篇文章主要内容时,就背离了综述文章的初衷,编辑在审稿时应加以辨析。

2.1.2 参考文献格式标注要正确

目前,大部分编辑对参考文献的校对主要集中在参考文献的格式核对,该部分工作量非常大,而且要求非常细致。订正标注不正确、不规范的参考文献要按照GB/T 7714—2005《文后参考文献著录规则》要求^[3],这里不作为重点阐述对象。

2.1.3 核查被引用参考文献是否真实存在,有无杜撰

对参考文献内容引用是否真实情况的检测一直是稿件修改过程中的盲点和难点。徐枫指出:“学术失范现象……表现为抄袭剽窃、一稿多投、虚假注释、不实参考文献、不当署名等。这些行为在学术期刊中之所以存在,有社会以及作者个人的因素,也与学术期刊自身在当今海量的学术信息中难以甄别判定有关”^[5]。的确,编辑很难做到对每条参考文献都逐一核对是否真实存在,主要是信息量实在是太大。有了学术不端文献检测系统,可以帮助编辑及时锁定有问题的参考文献并进行核查。

2.2 甄别参考文献是否真实的方法

学术不端文献检测系统设计的原理是对原封不动的复制粘贴抄袭进行比对识别,它能将待测文章中与文献库中被引用而没有标注的文献以红色标识出来。系统并没有专门对参考文献设检测项目,凭编辑的经验来说,凡是检测到的标注了参

考文献的重合部分(被标红的参考文献部分),都有参考文献引用方面的问题。^[6]参考文献不实引用要分几种情况。

2.2.1 检测到的标注参考文献的重合部分

有些作者在撰写论文时没有真正看过被引用的参考文献,或者从别人文章中看到就转引,并没有经过核实,这种情况经常导致引用参考文献张冠李戴。这就要求编辑对检测结果进行认真分析,凡是系统检测到的标注了参考文献的重合部分,就应初步认定有抄袭的嫌疑。

双击左键,页面上会出现对话框,框内有该段文字真正来源的显示,点击对话框就会打开被抄袭的文章,具体查看这段文字与被抄袭文字的重合情况。编辑可以根据被测者文后提供的参考文献,结合重复文字的出处进行判定,如果两者提供的参考文献出处不一致,则说明作者可能错误地引用了参考文献。但是,要注意的是,不能凭借系统标红就说作者有抄袭现象,笔者曾提出:系统收录不全、作者误引“抄袭文献”、引用了公有领域思想^[2]引起的标红现象要经过核实,然后再适度排除。

另外值得一提的是有小部分的参考文献作者已经标注,还是被系统认定为抄袭,而且被抄袭的文献与作者提供的是同一文献,这可能与该系统公布的查全率和查准率均>80%、并未达到100%有一定的关系,^[7]因此需要编辑要有足够的耐心认真核查。

2.2.2 检测到的未标注参考文献的重合部分

另外一种情况是没有标注参考文献的某段文字被系统标红,显示有抄袭情况发生。这种情况一方面提示作者引用了参考文献却没有标注,另一方面说明作者有抄袭现象。编辑可以根据文章是否有学术价值决定文章是否给作者继续修改的机会。

2.2.3 标注的参考文献未被检测出重合

这种情况并不代表该篇文章参考文献引用规范或者没有进行抄袭,有两种可能导致这一结果。

(1)作者规范地引用了参考文献。这种是比较理想的状况,碰到这样的文章,可以直接通过,送入下一个检测环节。

(2)作者引用的参考文献经过了改写。跟文章改写的道理一致,参考文献也有被作者有意引

用或改写的情况,改写的部分只要没有和系统内的文字重复,系统就认为作者没有抄袭,也就不会把这段文字标红,看起来作者引用的参考文献是正确的,其实情况不尽然。

笔者曾拿1篇未发表的500多字的读后感作为测试文章,第一,在文中选取8段话,在其后标注了8个参考文献,在选取的8段话中,其中一条是真正的引用,一条是经过稍微改写的引用,其余是随意标注;第二,在文后列出与其无关的5条参考文献和1条真正引用的参考文献。

检测的结果:6条标注的参考文献都顺利通过,1条原文引用的被标红,1条经过很小幅度改写的参考文献被检测出来,部分呈红色。根据检测系统的自适应多阶指纹(AMLP)特征检测算法比对原理^[8],可以初步断定,只要没有复制粘贴类型的抄袭,无论怎样标注参考文献都不会被检测出来。

对没有被系统检测出来的有抄袭嫌疑的参考文献,编辑可以掌握如下几条规律,如参考文献中提到了具体的作者名,可以到后面进行核对。另外作者引用的理论一定要是本学科最具代表性、最权威的言论。还有可利用学科领域进行辨别,被引用的部分论点与被引参考文献的主题是否相符。通过这样的辨别,基本可以排查大部分的抄袭。

2.3 对CNKI拓展参考文献查实功能的建议

2.3.1 建立参考文献真伪判定功能

检测系统应新增对文中标注参考文献字段是否来源真实的比对。

基于检测系统文字比对的功能,实现参考文献的查实在理论上可行。将待测的参考文献字段与待测文后提供的参考文献来源比对,若待测文章文后标注的来源文献中没有比对到该段文字,应予以绿色提示,显示该参考文献可能出处有误。

2.3.2 增加对单项标红文字字数和其占全文字数的百分比统计

在检测报告中,有专门一项是报告参考文献字数的项目,但是,这个是针对文后附注参考文献的字数,而不是文中引用的参考文献的字数。^[6]最终的抄袭数量和程度要靠编辑的后期分析判断。

系统可以增加单项标红文字的百分比和字数统计,以方便检测者在文章的总复制比中增减有

问题、或被误判的参考文献比重,用以实现在数量上判断是否符合10%的适度引用的标准。

3 结论

CNKI检测系统的最大功能是汇聚了海量的文献,可以在一定程度上和特定的范围内实现文章的比对功能。笔者认为除了在稿件初筛过程中能对抄袭状况进行把关,检测系统还在帮助检测者在遴选审稿专家和监测参考文献使用是否恰当方面具备开发价值。笔者建议经过系统功能适度调整,一是可以利用重合的途径检索出需要的专家,并建议建立可供所有编辑共享的专家库;二是可以拓展参考文献的查实功能,用以解决编校流程的难点,最终实现检测系统的管理机构、编辑以及净化学术环境的多赢局面。

参考文献

- [1] 赵蔚. 高校教育科学类学报栏目设置刍议[J]. 宁波大学学报, 2012(3): 64-67.
- [2] 赵蔚. 对学术不端检测结果的再研判标准初探——兼辨“文字复制比”与“文章抄袭率”[J]. 中国出版, 2011: 279(22): 20-23.
- [3] 陈浩元. 著录文后参考文献的规则及注意事项[J]. 编辑学报, 2005, 17(6): 413-415.
- [4] 中华人民共和国著作权法[M]. 注释本. 北京: 法律出版社, 2002.
- [5] 徐枫. 呼唤专家匿名审稿制——建立综合性学术期刊评价体系应重视的一个问题[J]. 中国出版, 2001(9): 45-46.
- [6] 吴均, 江润林, 张晓琴. 利用学术不端文献检测系统研究科技论文中存在的问题[J]. 中国科技期刊研究, 2010, 21(5): 636-639.
- [7] 中国学术期刊(光盘版)电子杂志社AMLC管理办公室. 系统简介[EB/OL]. (2008-12-26). [2012-10-20]. <http://check.cnki.net/amle/2008-12-26>.
- [8] CNKI科研诚信管理系统研究中心. 学位论文学术不端行为检测系统的研制说明[EB/OL]. (2010-08-31). [2012-10-20]. <http://check.cnki.net/smlc2/SimResult.aspx?Group=2010031613355304122745>. ■