

期刊学术不端文献检测系统误检分析*

谢文亮¹⁾ 李俊吉²⁾ 张宜军³⁾

收稿日期: 2013-03-08
修回日期: 2013-05-17

1) 《广东商学院学报》编辑部, 510320 广东省广州市赤沙路 21 号, E-mail: xwl@gdce.edu.cn
2) 太原科技大学 030024, 山西省太原市万柏林区窰流路 66 号, E-mail: lijunji04@163.com
3) 沈阳农业大学出版部, 110161 辽宁省沈阳市东陵路 120 号, E-mail: xbzyj@126.com

摘要 通过精心设计测试样本,对中国知网的学术不端检测系统进行多方面实验测试,分析检测结果,发现学术不端检测系统仅要求文后参考文献著录项中的题名(包括副标题)必须精确,题名中漏字、错字、多字、英文字母大小写状态错误、中英文字符切换状态错误等都会导致误检,其他的著录项如主要责任者、文献类型标志、出版物题名以及页码等录入错误将不会导致误检;另外,文后参考文献著录项前必须单独另起一行,并用“参考文献”标明,否则也会产生误检。研究结果说明:由于检测系统的误检而导致误检率极高,要求期刊编辑在检测文章时,必须对文后参考文献的格式和文章题名加以重视,尽量避免误检的情况发生。

关键词 学术不端文献检测系统 误检率 文后参考文献 文字复制比 非引用复制

1 引言

中国知网研发的“学术不端文献检测系统”(以下简称“检测系统”)给期刊编辑部的质量保证把了第一关,但也带来一些误检的情况发生。编辑在检测过程中经常会碰到以下情况:系统提示该段文字内容为非引用复制,通过系统提供的源文献作对照,文字内容也的确来自该源文献,然而,经过仔细阅读稿件原文,发现该段文字是作者引自系统中所提供的源文献,并且作者在正文中也作了声明,在句末用上标的文后参考文献序号“[J]”标明,在文章尾部也提供了该引文文献。通过对比两个参考文献,从整体来看,作者提供的参考文献与检测系统中提供的参照源文献是基本相同。然而,检测系统还是将该段文字判断为非引用复制。一些期刊编辑看到检测系统所显示的结果中文字非引用复制率比较高,就仓促将稿件定性为存在学术不端行为,直接退稿。这种情况造成了大量的好稿无辜被退掉,没有尽到期刊“挖掘人才,发现好稿”的责任。

鉴于此种情况发生,吉家友认为,检测系统初步的诊断不能作为结果,复制不能一律等同于抄袭^[1];赵茜认为在使用检测系统要注意一个“度”,人工和系统结合^[2];谭华,崔浩等认为编辑要全面分析稿件的内容、文中的参考文献的引用是否合理,避免因过分依赖检测系统审稿产生的审稿偏差^[3];赵蔚^[4]提出了要认真区别“文字复制比”与“文章抄袭率”。

然而,当前的研究很少通过实验技术手段,对检测结果

进行多方面试验,深层次研究误检的原理以及应对的策略。文献[2]中,赵茜发现在检测结果中文章提供的文献和系统给出的参考文献是一样,但系统还是将段文字判断为抄袭,经过对照观察,原文作者提供的文献中文献类型标明的是“[M]”,即书目,而检测系统中标明的文献类型是“[J]”,即期刊杂志,于是文献[2]就断言,检测系统以标明的参考文献类型作为判断标准。然而,笔者做过试验,将一篇文章中同一条参考文献的文献类型分别写为书目“[M]”和期刊“[J]”,在检测系统中检测,系统都识别该段文字为引用,并没有误断为非引用复制。因此,有必要对检测系统的误检作深入研究,探究检测系统误检的原因,并分析系统误检的严重程度。

2 检测系统误检实验与分析

学术论文撰写过程中经常要引用到其他公开发表文献的内容,必须在文后提供所引文字内容的出处。《文后参考文献著录规则》(GB/T 7714-2005)对期刊的析出文献著录制定了规范,其著录格式为:

析出文献主要责任者.析出文献题名[文献类型标志].连续出版物题名:其他题名信息,年,卷(期):页码[引用日期].获取和访问路径.

为了衡量检测系统的误检情况,笔者提出了误检率的计算公式:

误检率 = 未修改前非引用复制比 - 修改后的非引用复制比。

* 基金项目:中国高校科技期刊研究会基金资助课题(批准号:GBJXC1249)

2.1 文后参考文献误检为正文内容

一些期刊的文章结构编排规范中要求将“参考文献”这几种字也放进文章标题结构当中。笔者在检测一篇文章时,该文章的结构是:一、引言;二、样本的选取、变量描述;三、

实证研究;四、结论;五、参考文献。

检测该文章后发现参考文献的字体颜色大部分都是红色的,即检测系统将文后参考文献当成文章正文进行检测,如图1所示。



图1 文后参考文献被误当作正文内容处理

该文章的非引用文字复制比是 20.3% ,将“五、”去掉,即“参考文献”四个字独立出一行,重新检测后检测结果显示,文后参考文献内容变成浅灰色,非引用文字复制比变为 2.5%。误检率为: $20.3\% - 2.5\% = 17.8\%$ 。

分析:系统程序默认的情况下,文章的标题结构如“一、二、三、……”是属于文章的正文内容,如果用“五、参考文献”表示文后参考文献,则系统认为参考文献的内容也是正文,对其进行匹配检测。一旦发现有文字复制行为,无论正文中是否已经标明是引用,一律被认为是非引用复制行为。这种情况误检率比较大,需要特别小心对待检测结果。如果编辑不进行对照,只看到检测结果中非引用复制率高达 20.3% ,肯定作退稿处理。因此,笔者认为,除了要小心对待误检之外,还要严格实行在线稿件处理系统的“三审制”,将检测结果作为文档保存,以供下一阶段的审稿人审查。

2.2 文后参考文献著录项中的敏感与非敏感元素

选用另一篇文章作试验,系统检测结果显示文章的非引用复制比是 19.9%。通过检查生成的检测报告,发现文章中有一段文字已声明引用文后参考文献 15,而参考文献 15 与检测系统中提供的文献是一字不差。如:

[15]张文勤,石金涛等. 团队成员创新行为的两层影响因素:个人目标取向与团队创新气氛 [J]. *南开管理评论*, 2010(6): 22-30.

但检测报告依然将该段文字判断为非引用复制。笔者对该参考文献著录作了如下的试验:

- (1) 将文后参考文献中的序号 “[15]” 去掉;
- (2) 将主要责任者去掉;
- (3) 将期刊题名故意写错一个字,如“南开管理评论”写成“南开管理评”;

(4) 将出版日期作了修改,如“2010(6)”写成“2010(13)”;

(5) 将页码“22-30”去掉;

(6) 将文献类型 “[J]” 改成 “[M]”

(7) 将正文中的文后析出文献序号 “[15]” 分别放在句前和句后;

结果与分析:根据(1)-(6)所列的试验项对文后参考文献 15 分别作了修改,根据(7)对正文内容作了修改,上传到检测系统检测,结果非引用文字复制比依然是 19.9%。这说明检测系统中对文后参考文献中以上各项不敏感。

(8) 将文后析出文献标题中的中文冒号 “:” 切换成英文状态下的 “:”,

结果与分析:根据(8)所列的项目对文后参考文献 15 进行修改,检测结果从 19.9% 的非引用文字复制比降低为 13% ,系统的误检率是 6.9% ,说明检测系统对标题中的符号敏感,如中英文切换状态、英文字母大小写等。

(9) 将析出文献标题中的“团队成员创新行为的两层影响因素:个人目标取向与团队创新气氛”去掉一个“氛”字;

结果与分析:检测系统显示非引用文字复制比由 13% 重新回升到 19.9%。这说明检测系统对析出文献标题中的文字敏感,文献标题中少一个字,多一个字或错一个字都会报错,系统认为文章中提供的文后参考文献与系统数据库中的源参考文献不相符,进而将该引用的文字判断为非引用复制。

(10) 文后参考文献题名缺少文章副标题;

结果与分析:在检测另一篇文章时发现,该文章的非引用文字复制比是 5% ,从正文的内容来看,是属于引用复制,并且作者提供的参考文献题名与检测系统中参考文献正标题同。如:

[7]纪宣明,黎丽.东西部上市公司经营绩效的实证分析与评价[J].金融研究,2004,3:109-115

但系统中将引用该文献的文字认为是非引用复制。经过对比,发现检测系统中所提供的源参考文献中副标题是存在的,但在检测的论文中没有提供副标题。对文后该参考文献内容稍作修改,加上副标题,如:

[7]纪宣明,黎丽.东西部上市公司经营绩效的实证分析与评价——以京、蜀两地上市公司为例[J].金融研究,2004,3:109-115

重新检测后结果表明:非引用文字复制比从5%低到了0.5%,误断率是4.5%。

这说明检测系统对副标题敏感,副标题必须准确加入文后参考文献著录中。

3 总结

通过以上的测试实验和分析,总结出中国知网检测原理:

第一 将要检测的文章分为正文和参考文献,如果文章结构中没将“参考文献”在文后单独列开,检测系统将文后参考文献当作正文检测;

第二 将正文与数据库进行匹配检测;

第三 如果正文有文字复制情况发生,则提取数据库中

源文件的参考文献的题名,题名包括主标题和副标题;

第四 检测系统将所提取的题名与所检测文章的文后参考文献题名作精确比较,当且仅当一一对应,检测系统才将该文字复制标为引用,否则检测系统认为是非引用文字复制,即对参考文献著录的标题零容错,对其他项宽松。

至于文后参考文献中的其他著录项,如主要责任者、连续出版物题名、页码等,检测系统并没有对其进行检测,因此对其他项的错误并不敏感。在正文中,无论参考文献序号“[1]”、“[2]”等有没标出,检测系统对此不敏感。

结合以上的分析,总结出学术不端检测系统产生误检的可能原因如下:

(1) 文章作者在文后参考文献著录时没注意规范,将题名写错;

(2) 来源期刊缺乏统一规范,像文中所举的例子,有些期刊将冒号写成中文字符“:”,也有期刊写成英文字符“:”,结果导致读者无法分辨出中英文字符切换状态下正确的字符,因此在文后参考文献著录时发生错误。

了解以上各种误检情况以及其产生的可能原因后,期刊编辑在对作者来稿进行检测时,需要仔细和谨慎,如果有非引用文字复制产生时,必须生成检测报告单,并且为了方便对照,在生成报告单时,选中“全文对比”选项,如图2所示。

2012-t-000323_editing[1] 总文字复制比: 13.2% (1715) 总字数: 13011		
1	<ul style="list-style-type: none"> 基于因子分析法的我国农业上市公司经营绩效评价 彭源波, - 《生产力研究》 - 2012-01-15 	8.5% 是否引用: 是
2	<ul style="list-style-type: none"> 公用事业类上市公司经营绩效的实证分析与评价 纪宣明,陈似海, - 《宏观经济研究》 - 2004-08-20 	4.2% 是否引用: 否
3	<ul style="list-style-type: none"> 东西部上市公司经营绩效的实证分析与评价——以京、蜀两地上市公司为例 纪宣明,黎丽, - 《金融研究》 - 2004-03-30 	3.9% 是否引用: 否

原文内容	相似内容来源
<p>此处有 36 字相似</p> <p>因此,研究粮油上市公司的经营绩效,对于提高农业类上市公司的生产经营质量,</p>	<p>基于因子分析法的我国农业上市公司经营绩效评价 彭源波,《生产力研究》-2012-01-15</p> <p>1. 因此,研究农业上市公司的发展情况,对于提高农业上市公司的经营绩效。</p>
<p>此处有 88 字相似</p> <p>效益,具有十分重要的意义。本文首先运用12个财务指标对我国粮油上市公司2007-2011年的经营绩效进行总体评价,然后利用因子分析和聚类分析方法对同期各家公司经营绩效进行比较,</p>	<p>基于因子分析法的我国农业上市公司经营绩效评价 彭源波,《生产力研究》-2012-01-15</p> <p>1. 意义。本文先利用12个财务指标对我国农业上市公司2002—2006年的经营绩效进行总体评价,然后利用因子分析法对同期的经营绩效进行</p>
<p>此处有 161 字相似</p> <p>2000,最早将因子分析引入对农业类上市公司经营绩效的研究。他们通过考察这些公司的盈利能力、偿债能力、营运能力、成长能力和股本扩张能力等,综合评价了农业上市公司的经营状况。朱丽莉和王怀明(2004)利用因子分析法对2001-2003年27家上市公司的绩效进行了实证分析,得出其总体经营绩效不佳的结论。林乐芬(2004)运用</p>	<p>基于因子分析法的我国农业上市公司经营绩效评价 彭源波,《生产力研究》-2012-01-15</p> <p>1. 2000年将因子分析法引入对农业上市公司经营绩效的分析,通过考察农业上市公司盈利能力、偿债能力、营运能力、成长能力和股本扩张能力五个方面,综合评价了农业上市公司财务状况。朱丽莉、王怀明(2004)利用因子分析法对2001—2003年27家上市公司的绩效进行了实证分析,得出总体经营绩效并不理想的结论。林乐芬(2004)运用</p>

图2 “全文对比”检测生成报告

观察检测报告右边中提供的参考文献著录情况,再对比所检测文章中文后参考文献提供的著录项中的文章题名,如果基本一致,但系统对该段文字判断为非引用复制,则须:

- (1) 注意标题(包括副标题)是否少字、多字或错字;
 - (2) 如果标题中有英文字母存在,注意字母的大小写;
 - (3) 如果标题中有数字和符号,注意字符的中英文切换状态;
 - (4) “参考文献”四个字是否在文后单独另起一行。
- 然后根据上面分析的结果修改参考文献后再检测,则可

得到相对客观的、误检率低的检测结果。

参考文献

- 1 吉家友. 学术不端文献检测系统数据分析. 中国出版, 2010, (4): 27-29
- 2 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”. 中国出版, 2010, (10): 59-61
- 3 谭华, 崔洁. 学术不端文献检测系统的使用建议. 编辑学报, 2010, 22(2): 153-155
- 4 赵蔚. 学术不端检测结果的修正标准初探——基于“文字复制比”与“文章抄袭率”的辨析. 中国出版, 2011, (22): 20-23

引用矮化 ——论参考文献的不当引用*

闻浩鲁立**

收稿日期: 2013-05-30
修回日期: 2013-07-30

南京军区南京总医院《医学研究生学报》编辑部 210002 南京市中山东路 305 号 A7 信箱, E-mail: 62840331@sohu.com

摘要 参考文献是学术论文重要的组成部分。科学合理的引用文献对提升论文学术质量具有重要意义。但放弃了文献引用的必要性与合理性,盲目追求源文献的高水平与权威性易造成文章结构混乱、主题偏移,进而影响论文整体性表达,严重降低文献的引用效果,形成论文的“引用矮化”现象。文中就学术论文文献“引用矮化”现象的成因、不良影响与预防措施作一综述。

关键词 引用矮化 不当引用 参考文献 学术论文

参考文献是学术论文重要的组成部分,是科学研究线索的体现形式。标注参考文献的意义在于体现论文的学术水平、科学依据的背景、起点及深度,作者对他人劳动的尊重及对科学研究的严肃性、相关科研内容的延伸性^[1]。以往有研究针对学术文献引用的格式、作用与意义展开讨论^[2],但较少关注“经典文献”不合理引用对论文学术价值与可读性产生的不良影响。现实中,少数缺乏科学精神及学术严谨性的作者在将高水平文献引用并混搭后,却没有体现出更科学的学术思想及更高的学术水平,甚至有误导读者之嫌;且文献的牵强引用造成篇章结构的臃肿,思想内容也不及参考文献本身,反而降低了文献组合后的引用效果,形成了文献的“引用矮化”现象,即引用意义的“1+1<2”。文中就参考文献“引用矮化”现象的形成原因、对论文学术价值的影响以及预

防措施作一综述。

1 “引用矮化”现象形成原因

“引用矮化”现象是由于作者在引用文献过程中忽视了对文献必要性与合理性原则的把握,所引文献对文章主题及结构没有产生积极作用。形成论文参考文献“引用矮化”现象主要来自于作者与编者两方面的原因。

1.1 作者原因

1.1.1 缺乏正确引文认识

部分作者缺乏对正确引用文献的认识,论文写作过程中只注重对研究内容、方法与结果进行阐述,在需要文献佐证

* 基金项目:南京军区南京总医院院管课题(2013050)

** 通信作者:鲁立