

# 正确看待“科技期刊学术不端文献检测系统”检测结果

孔琪颖 蔡 斐 张利平 徐 晓

(《航空学报》杂志社,100191,北京)

**摘 要** 归纳影响“科技期刊学术不端文献检测系统”检测结果的要素,对初审中运用该系统检测出的不同等级文字复制比的文章的学术不端可能性进行了分析。结合实际检测案例,得出该系统能有效帮助初审,但尚不能取代初审的结论。

**关键词** 学术不端文献检测系统;初审;编辑

**Analysis of AMLC check results** // KONG Qiyong, CAI Fei, ZHANG Liping, XU Xiao

**Abstract** Main factors that affect the results of Academic Misconduct Literature Check (AMLC) are summarized. Then, whether a manuscript has piratical content, and which types of plagiarism the manuscript may have are analyzed combining with the editors' initial review experience. Analyses show that the AMLC can really facilitate the initial review by the editorial office, but AMLC check results can not replace the initial review.

**Key words** AMLC; initial review; editor

**Author's address** Press of Chinese Journal of Aeronautics, 100191, Beijing, China

近年来,学术不端问题日趋严重,学术诚信问题已经引起政府和管理部门的高度重视。正如2009年3月15日,教育部长周济在高校学术风气建设座谈会上强调的——对学术不端行为要“零容忍”<sup>[1]</sup>,浙江大学何海波学术不端事件相关责任人受到了中国高校和研究机构对类似论文造假行为最严厉的一次惩罚<sup>[2]</sup>。

科技期刊编辑部作为科技论文的第一道防线,如何在初审中过滤品质不佳,特别是有抄袭等恶劣学术不端行为的稿件,也引起了编辑工作者的广泛探讨<sup>[3-7]</sup>。王淑华<sup>[3]</sup>于2001年提出利用中国期刊网初审的方法。随着网络技术的迅猛发展,近年来各编辑

部初审时也切实实践类似的网络审稿方法<sup>[6-7]</sup>;然而,当文章数量成为学术成果的考核重要标准之一时,投稿量明显增加,人工逐篇进行数据库检索的工作量随之大大增加,并且难以发现多源抄袭等复杂的学术不端问题。“科技期刊学术不端文献检测系统”(AMLC)的面市无疑给这项工作带来了极大方便,很多编辑部开始使用该系统协助编辑人员对稿件进行初审。

AMLC是中国知网(CNKI)历经3年的研发成果,它以《中国学术文献网络出版总库》为全文比对数据库,可检测抄袭与剽窃、伪造、篡改、不当署名、一稿多投等学术不端文献,并支持各刊自建比对库<sup>[8]</sup>。

学术不端行为的认定是一个严肃的问题,这涉及举报人、被举报人以及科研基金资助单位的荣誉、利益等方方面面的问题;因此,对于不端行为的认定需要制订规范的程序,需要保证程序的公正<sup>[9]</sup>。2009年新疆部分高校引进AMLC的姊妹版本“学位论文学术不端行为检测系统”,用以检测毕业论文引发的争论<sup>[10]</sup>,也提醒科技期刊编辑在使用AMLC时应正确看待系统的检测结果。

## 1 影响AMLC检测结果的要素

分析AMLC的工作流程可以看出影响检测结果的因素如下。

1)待检测文章。

①作者相同与否,直接决定有文字重合度的稿件是自抄、不当署名,还是抄袭,问题严重性自然不同。

献引用的权威性原则,其根本目的是要重视文献引证的质量、可信度和实际效果。作者要端正引文动机,要在广泛收集相关文献的基础上,尽可能选择有必要引用的权威文献加以引证;编辑和审稿人则关键要看引文的实际权威性,而不是文献作者或期刊的权威性,并要认真分析引用文献在论著中的学术论证功能,以更好地鉴别论文的质量和发表价值。

## 5 参考文献

[1] 陶范.参考文献引用原则辨析[J].编辑学报,2006,18(4):252-254

[2] 陶范.参考文献的引用应遵循代表性原则[J].编辑学报,2008,20(6):483-484  
[3] 朱大明.简析学术论文引言中的参考文献[J].编辑学报,2005,17(3):190-191  
[4] 费多益.科学价值论[M].昆明:云南人民出版社,2005:28-35  
[5] 王小唯,吕学梅,杨波,等.学术论文引言的结构模型化研究[J].编辑学报,2003,15(4):247-248  
[6] 朱大明.应注重参考文献引用的学术论证功能[J].科技与出版,2008(12):43-45

(2009-06-11 收稿;2009-07-13 修回)

②文章的可检测度。这里指文章文字居多,还是公式、图表居多,汉字多,还是外文多。由于目前 AMLC 还不能有效检测公式、图表、外文,所以可检测部分的比例不同,检测结果与真实相似度的接近程度也不同。

③文章的格式。虽然系统支持 .pdf、.caj、.doc、.txt 等多种格式,但是,实际使用中发现,有时对不同格式的同篇文章处理结果并不相同。

2) 比对数据库。这是 AMLC 的基础与核心。完善的比对数据库必须全面、完整,及时更新。

3) 检测指标。AMLC 提供多种检测指标,包括文字复制比、最大段长、首部重合度、尾部重合度等。其中最直观的指标是文字复制比,这也可以算作系统提供的主要检测结果之一。

## 2 AMLC 检测结果分析

下面对工作中使用 AMLC 辅助审稿时遇到的实际情况进行具体讨论,仅考虑研究内容已经符合刊物刊载范围的稿件,讨论的抄袭嫌疑不区分是抄袭别人还是自抄,并假设投稿承诺书中作者承诺稿件中已经正确标引他人工作、本稿相关内容也从未发表及已投他处。

考虑到在我国一般认为相似度 >30% 的稿件有抄袭嫌疑,而 >50% 的稿件有全篇抄袭的嫌疑;所以,相似度在 30% ~ 50% 范围内的稿件会被重点排查,而相似度 <30% 的稿件容易被直接通过, >50% 的稿件又容易被直接拒稿。但实际情况比较复杂,不宜简单以检测数据来判定稿件抄袭与否。

**2.1 文字复制比较低的稿件** 是否文字复制比 <30%,甚至标准更严格些, <25% 的稿件就可以通过初审呢?

根据第 1 章的分析,相似度不高仅能代表比对数据库中沒有与其文字重合度高的文章,可能的情况如下。

1) 确实为具有独创性的好文章,可以通过初审。

2) 没有直接抄袭他人语句,但也没有创新。比如文章内容为现有方法的直接应用,或者属于思想性抄袭,抄袭现象被很好地掩藏。对于简单的成熟方法直接应用型稿件,编辑根据自己的知识积累往往能发现其中的一部分,可以将其退稿。

3) 文章中不可检测部件(图、表、公式、外文)较多,导致检测结果与真实结果相差较大。如果是抄袭公式或者图表,只要被检测文章文字复制比处不是显示绿色条(即文字复制比 <5%),那么下载、查看文字重合来源的文章也很容易被发现,特别是,系统往往会

在公式前的引导段落、图题或表题检测出重合文字。编辑根据问题的严重程度也能作出正确的处理。而对于通过翻译抄袭另外一种语言文章的稿件,多数编辑恐怕力不从心,无奈让其蒙混过关。

4) 抄袭比对数据库中尚未记录的文章。可细分为以下情况。

①在网络上通过别的途径,比如别的网络数据库、论坛、博客等传播的文献。虽然《中国学术文献网络出版总库》收录了 1915 年以来的超过 5 000 万条的数据,但仍未涵盖所有网络上传播的电子数据,这时就需要在百度、google 等搜索引擎以及万方、维普等其他数据库中再次检索作者、题名、关键词等信息元,进行深度排查。

②本编辑部的退稿、在审稿或已录用待发表的稿件,甚至包括部分刚出版还未来得及加入《中国学术文献网络出版总库》的稿件。这就要求编辑部及时更新个人对比库,保证其中尽可能收录本刊尚未公开的文献,特别是多产作者(课题组)的稿件。

③其他编辑部未公开的稿件,特别是在审及已录用待发表的稿件。由于不同刊物间的信息难以实现共享,而且很多期刊发表周期长达 1 年,这使一稿多投甚至一稿多发有机可乘。

④其他未在网络上公开的文献,如仅有光盘版本或纸介质版本的文献等。这类稿件在初审时也较难发现问题。

**2.2 文字复制比较高的稿件** 同样,对文字复制比 >50% 甚至更高些的稿件也应根据具体情况进行具体分析。

1) 确属全篇抄袭,毫无疑问可以退稿。

2) 退稿重投,且先前被退稿件被加入了比对数据库。此时,通过系统给出的 2 版稿件的异同比对,再结合退稿意见,如果新投稿未能妥当解决原审稿意见提出的问题,也可直接退稿。

3) 同一作者撰写的有延续性的研究成果。对于这类稿件,笔者认为不适合直接退稿,具体分为以下 2 种情况。

①将一篇稿件拆分发表,单纯追求数量的稿件拆分可能造成几篇稿件的信息量都不充分,此时可以先与作者联系,告知利弊,建议其修改合并,再作出退稿与否的决定。

②在我国目前的科研评价体系下,劝说作者将多篇独立成文的稿件合并为一篇长文是非常困难的。如若不能合并,则应仔细检查涉及的稿件:引言和结论是否是针对各篇文章的具体研究内容来陈述的;上一篇文章的研究内容在后续稿件中是否是简单扼要地引入

而非大段或整篇搬抄。如果作者未能有针对性地突出各篇文章的创新点,则应退回作者修改。

4)与同一作者的学位论文相似度高。此时需结合仅检测出与该学位论文相似,还是同时与别的文献也相似,来作初审决定。

5)综述类文章。由于好的综述类文章肯定有体现作者独到见解的内容,如果检测结果显示稿件完全是他人成果的简单堆砌,那么其价值是值得商榷的。

文字复制比在30%~50%范围内的稿件可以按照2.1节及2.2节的分析进行类似处理,不再赘述。

**2.3 文章格式对检测结果的影响** 使用中发现,虽然AMLC支持多种格式的文章比对,但实际比对效果却可能不同。

使用过该系统的编辑同人可能都有过成功提交PDF文档后却显示全文太短的体会,由于系统本身的技术问题,实际未能成功上传稿件,这就是文章格式影响的一个方面。

笔者初审稿件时,在中国期刊网中检索其一篇参考文献,无意中检索出2篇作者不同而题名一样的文章,将这2篇文章的.caj版本下载下来并利用AMLC检测。在具体查看重合文字来源列表时,发现2篇稿件之间相互无重合,然而检测结果却也意外地显示2篇文章与其自身的对比结果也并非100%(一篇是64%,另一篇是77%)。

下面这个例子,能更直观地反映出这个问题。《航空学报》个人比对库中的一篇文章(.doc格式,为作者提交的录用定稿)已经在中国期刊网上发布(具体采用的比对文件格式未知,发布前对录用定稿进行了编辑加工但无大修改),新提交的一篇稿件被检测出与这2篇文章有文字重合的问题;但检测结果中该稿与个人对比库中的稿件间的文字复制比为46%,而与同一篇文章在中国期刊网中的版本的文字复制比为26%,不同的格式竟然造成检测结果间20个百分点的差别。

检测系统不管有多先进,程序有多优良,在它们刻板地按照指令飞速执行时,却终究因为无法独立思考,而可能得出不那么正确的结果。

### 3 结束语

AMLC是智能很强的系统,能帮助编辑更快速、更有效地初审稿件;但是,稿件的情况非常复杂,应该认真对待每篇稿件,对具体情况具体分析,特别是学术不端问题的认定是个敏感问题,所以在使用AMLC过程中,应牢记它只是一个机器,所产生的结果报告只能用作审稿时的参考,而不应取代编辑在初审中的主观能动性。

使用AMLC时要注意完善比对库,以便其更好地为我们服务;同时,仅依赖AMLC进行初审并不够,依然要注意结合使用多种检测手段。编辑也应注意增加自己的专业知识积累,不断提高鉴别稿件的能力。

### 4 参考文献

- [1] 教育部部长强调对学术不端行为“零容忍”[EB/OL]. (2009-03-16)[2009-05-15]. <http://edu.sina.com.cn/l/2009-03-16/1126167330.shtml>
- [2] 中国重拳打击学术不端行为[EB/OL]. (2009-03-17)[2009-05-15]. <http://tech.sina.com.cn/d/2009-03-17/10592916048.shtml>
- [3] 王淑华. 如何利用中国期刊网审稿[J]. 编辑学报, 2001, 13(4): 225-226
- [4] 郑淑芳,徐前进,彭金旺. 科技论文初审评价方法分析[J]. 武汉科技大学学报:社会科学版, 2007, 9(5): 542-544
- [5] 何洪英,李家林,朱丹,等. 论科技学术期刊论文的编辑初审[J]. 编辑学报, 2007, 19(1): 17-19
- [6] 张建合. 学术期刊编辑初审探微[J]. 编辑学报, 2003, 15(1): 23-25
- [7] 周英智,李伟,孙瑶,等. 初审中利用数据库检索论文抄袭行为[J]. 编辑学报, 2008, 20(4): 319-320
- [8] 科技期刊学术不端文献检测系统使用说明书[EB/OL]. [2009-05-15]. <http://check4.cnki.net/amlc/help>
- [9] 孟伟. 西方发达国家如何应对科研不端行为[J]. 科技导报, 2006, 24(8): 91-94
- [10] 新疆高校使用反剽窃软件治理学术不端遭质疑[EB/OL]. (2009-04-22)[2009-05-15]. <http://news.sina.com.cn/c/2009-04-22/040917659440.shtml>  
(2009-05-22 收稿;2009-07-21 修回)

### 2008年新闻出版学科高被引作者(总被引频次/被引文章数)

- |                |               |               |               |                |
|----------------|---------------|---------------|---------------|----------------|
| 1. 陈力丹(90/50)  | 2. 蔡雯(77/23)  | 3. 游苏宁(64/11) | 4. 喻国明(45/26) | 5. 朱大明(40/15)  |
| 6. 李武(36/4)    | 7. 郑保卫(30/18) | 8. 陈浩元(29/2)  | 9. 宫福满(27/15) | 10. 邵培仁(27/12) |
| 11. 彭兰(27/7)   | 12. 姚远(26/9)  | 13. 陶范(24/10) | 14. 董天策(24/9) | 15. 蒋永福(23/4)  |
| 16. 丁柏铨(20/14) | 17. 高钢(20/7)  | 18. 段京肃(20/5) | 19. 蔡骥(19/7)  | 20. 李良荣(19/6)  |

(林彤摘自《2009年版中国期刊高被引指数》)