

利用技术手段遏制科研不端行为的实践与思考

王胜海^{1,2} 钟 瑛^{1,2}

(1. 中国科学技术信息研究所, 北京 100038; 2. 北京万方数据股份有限公司, 北京 100038)

摘 要: 通过分析国内外论文相似性检测系统的特点、应用情况和相关技术的发展趋势, 指出利用技术手段遏制学术界的抄袭剽窃等科研不端行为问题, 不仅效果明显、威慑力大, 而且有利于引导学术规范、保证已发表科研成果的质量和促进知识产权保护工作。

关键词: 论文相似性检测; 抄袭; 剽窃; 科研不端行为

中图分类号: G

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2010.05.003

Practice of and Thinking about Using Technical Measures to Curb Research Misconduct

Wang Shenghai^{1,2}, Zhong Ying^{1,2}

(1. Institute of Scientific & Technological Information of China, Beijing 100038;

2. Wanfangdata CO., Ltd, Beijing 100038)

Abstract: By analyzing the characteristics, applications and related technology development trend of the domestic and foreign article similarity detection system, this paper shows that using technical measures to curb academic plagiarism and other academic misconduct is not only effective and deterrent, but also helpful to guide academic standards, ensure the quality of published scientific research achievements and promote intellectual property protection.

Keywords: article similarity detection, plagiarism, research misconduct

1 引 言

长期以来, 造假、抄袭、剽窃等科研不端行为在国内外学术界时有发生。世界最大的学术团体——美国科学促进会(AAAS)曾对其会员做过一次问卷调查, 结果显示48%的被调查者有捏造、篡改、剽窃等行为^[1]。我国近年来也有一些学者被曝光涉嫌抄袭剽窃行为, 其中被查实的北京大学王铭铭教授、西南交通大学黄庆教授等人都受到了一定形式的处理。科研不端行为不仅阻

碍并影响了科学研究的正常进行, 损害科学共同体的尊严, 也破坏了学术发展的生态环境^[2]。世界上有大规模科技投入的国家, 都十分重视遏制科研不端行为, 并采取了制度规范、宣传教育和监督约束等多方面的对策。随着数字化时代的到来, 技术手段被用来进行造假和剽窃, 但技术手段同时也是遏制科研不端行为的实用工具和有效的威慑力量。笔者认为, 通过完善反抄袭剽窃技术和其他相关检测技术, 进而扩大技术手段在学术出版和科研项目立项与成果验收中的应用, 将提高我国科研水平和科研成果质量。

第一作者简介: 王海胜(1975—), 男, 高级工程师, 研究方向: 数字图书馆。

收稿日期: 2010年8月30日。

2 文档相似性检测系统的概况

十多年来,国内外一些研究机构、企业和学者陆续研发出采用相同或不同的技术、面向不同行业与用户的多种剽窃检测系统,或称“文档相似性检测系统”。其中应用比较广泛、知名度较高的有如下一些系统。

2.1 Turnitin 和 CrossCheck 系统

美国 iParadigms 公司开发的这两个系统通过与主要国际学术出版集团签订协议,将内容丰富的学术数据库作为比对资源,其他比对资源还包括互联网网页和学生提交的论文等。教师利用 Turnitin 系统可以检查学生作业的独创性和是否存在剽窃问题。据称,全球 110 个国家的近万所中学和大学已成为该系统的用户^[3]。而由电子学术出版商协会 CrossRef 与 iParadigms 公司联合开发 CrossCheck 系统,主要供学术期刊检测投稿中可能存在的剽窃问题,其用户包括自然出版集团、爱思唯尔、施普林格和英国医学期刊出版集团等。

2.2 eTBLAST 文字相似性搜索引擎系统

该系统是基于网络的相似文本搜索引擎,由美国弗吉尼亚生物信息研究所创新实验室提供。用户通过比较投稿论文摘要和生物医学等方面学术数据库的相关信息,可以检测出涉嫌重复发表和抄袭剽窃的论文。用于比对的信息资源包括美国医学索引(MEDLINE)和美国宇航局数据库等海量数据库的内容。

2.3 论文相似性检测服务(PSDS)

该系统由北京万方数据股份有限公司开发,是基于中国期刊论文数据库和中国学位论文数据库中的文献,面向机构用户和公众提供论文比对分析服务。该系统操作便捷,学生和科研工作者都能够以较低的成本进行自主查询。该系统个人用户的日均检测量上千篇,团体用户增长迅速。

2.4 学术不端文献检测系统(AMLC)

该系统由同方知网开发,以“中国知识资源总库”为全文比对数据库,可检测出涉嫌抄袭剽窃、一稿多投等学术不端行为的文献,其用户主要包括国内学术期刊编辑部、研究生培养单位、出版社、科研管理部门和大学行政部门等。AMLC

系统主要面向机构用户。自 2008 年底推出以来,在国内高校研究生学位论文检测方面得到广泛应用,并在学界引起很大反响。

2.5 ROST 系统

该系统^[4]是由武汉大学沈阳副教授研制的,是一款由学者个人研发的开放源代码软件,用户可以在开发者的个人博客上免费下载。该系统的检测范围以网络信息资源为主,主要用于少量论文检测以及改善期刊编辑部的业务流程。

以上大多数检测系统都在进一步开发与完善中。例如,大多数系统都在不断增加用于比对的信息资源;在检测的可靠性方面进行新的改进,使检测结果不受作者修改个别词句或调整部分段落的干扰,并实现不同语种文献之间的比对等^[5-6]。

3 问题论文的主要特征

根据北京万方数据股份有限公司利用其知识服务平台对上千篇期刊论文和学位论文进行检测的结果,发现其中涉嫌抄袭剽窃、重复发表等问题论文具有以下主要特征。

(1) 显著的抄袭。作者引用他人作品的长度大致属于合理引用,但并未注明自己是在引用他人作品。也有些论文中的大部分内容都可以找到相似度极高的文献,即作者整段甚至整篇地抄袭他人的作品。

(2) 不当引用或引用不规范。最常见的是作者在引用文献时采取转引的方式,导致引文内容并不相关,或重复原来引文及参考文献出处中存在的错误。虽然引用原始论著并不是法律要求,但却是学术界的一种道德要求。美国化学会(ACS)曾在其《作者的道德责任》中明确指出,“作者有履行检索并引用(与本人工作)密切相关的原始论著的责任”。

(3) 重复发表。有些论文曾在多个期刊上发表,其中一些论文在发表时变换了论文标题或作者署名顺序以进行一定的伪装。由于编辑人员把关不严,也有个别论文在同一期刊的不同刊期中发表的情况。

(4) 自我抄袭。论文作者的自我抄袭也被称

为“高度自引”。不少作者出于增加发表论文数量和提高被引频次的目的，套用或大量引用已经发表文章中的内容，导致产生大量高相似度的文献。虽然这样的论文可能反映了作者学术思想的深化与发展，故无法判定为自我抄袭，但是如果缺少原创内容，只是补充修改原来论文中的核心内容，则难以避免拼凑的嫌疑。

以上这些行为可以给作者带来一定的学术资本和其他利益，但是对学术期刊而言，浪费了宝贵的出版资源；对读者而言，影响了他们获取更多的原创性知识；对作者所在单位和资助机构而言，可能在考核评价及项目资助等工作中被误导。

4 检测软件工具应用的效果

对于反抄袭剽窃软件的开发与应用，目前国内还存在不同的看法。例如，有些作者或编辑人员对软件本身的可信性持怀疑态度，认为在学术研究中借鉴他人成果是正当且不可或缺的，因此以文档相似度判断抄袭剽窃行为并不适当。也有些人认为我国出现的科研不端行为只是个别现象，如果真正需要在大学和期刊编辑部普遍使用“学术测谎仪”，可以说是学术界的悲哀。尽管有这样不同的声音，但从国内外现有检测软件的应用效果和技术发展趋势看，其所能发挥的积极作用是显而易见的。

4.1 有利于改善科研管理工作

美国卫生与公共服务部研究诚信办公室长期以来将抄袭剽窃检测与图片鉴定软件用于对科研不端行为的预防与调查工作，并为教育科研机构提供技术支持。美国国家科学基金会总监察长办公室也曾尝试使用 Turnitin 系统检测资助项目发表的成果，结果发现存在剽窃问题论文的数量大幅增加。我国科技部万钢部长在“科研诚信与学风建设座谈会”上的发言中表示，科技部将在项目申报和成果验收中使用抄袭剽窃检测软件等技术手段，主动排查科研不端行为^[7]。

4.2 发挥宣传教育作用

利用技术手段进行论文相似性检测，实际上并没有对科研人员或学生提出新的要求或增加新的压力，而只是增强他们的学术规范意识，鼓

励他们在科研活动中自觉遵循既定的标准和规范^[8]。在上述北京万方数据股份有限公司进行的论文相似性检测中，发现不当引用或引用不规范的情况占相当大的比例，这也说明加强学术规范方面教育与训练的重要性。Turnitin 系统的应用效果也说明，检测软件能够帮助教师指出学生作业中存在的恰当的引用或潜在的剽窃行为，并指导他们什么是正确的引用方式，而每份证明独创性的检测报告都是在鼓励学生正确地进行研究和写作^[9]。

4.3 为期刊编辑人员提供实用的辅助工具

在抄袭剽窃检测软件出现之前，国内外许多负责任的学术期刊如果怀疑投稿属于一稿多投或存在剽窃行为，会采取人工搜索和比对的方式，或凭借审稿人的经验进行验证和判断。反抄袭剽窃软件的开发应用，使编辑部的投稿检测方式发生了革命性的变化。尽管最后还是需要编辑人员和审稿人对论文是否确实存在剽窃问题进行定性，但无论是对可能存在问题论文进行筛查还是对论文是否存在剽窃问题的判断，技术手段都发挥了重要作用。

4.4 促进提高研究生培养质量

国外的大学利用 Turnitin 抄袭剽窃检测系统，明显减少了学生互相抄作业的情况。学术不端文献检测系统（AMLC）在我国研究生培养单位的使用，则大大减少了学位论文中的抄袭剽窃现象，而这样的效果和影响通过一般性的宣传教育是很难达到的。

4.5 促进保护知识产权

利用反抄袭剽窃软件，还可以帮助作者或学术出版机构发现自己拥有版权作品的潜在的抄袭者，从而维护自己的知识产权。有学者曾利用 Turnitin 系统的功能，找出自己文章的抄袭者和有关抄袭的证据^[10]。国际学术数据库也可以利用相关的技术手段维护学术期刊和作者的知识产权。

5 数据库服务商的作用

数据库服务商不仅掌握着大量的学术信息资源，有条件开发或购买检测科研成果的技术手段，而且与教育科研机构和广大科技工作者有着

密切的联系,因而在促进科研诚信方面能够发挥重要作用。

5.1 不断开发利用适当的软件工具

目前,现有的反抄袭剽窃软件还普遍存在一些不足之处。例如,文档相似性检测工具一般只能对论文的文本部分进行有效的检测,而对于论文中是否存在图片、表格造假和研究思路、公式抄袭剽窃等问题,只有依靠其他方式进行分析判断。现有检测工具的主要开发机构也在进行论文原创性检验等功能的开发。而对于图片造假或对图片进行不当的加工处理等问题,美国卫生与公众服务部研究诚信办公室很早便采用相关的检验方法和技术进行分析识别,如针对图表抄袭或将原图片进行颠倒、旋转、改动等行为,可以通过检查计算系列文件、检查底片被扫描的痕迹,使用滤镜进行分色、浮雕处理,检查对比度、质地、墨迹以及利用渐变映射等方式进行复原。实践证明,采用数字技术造假的行为,如利用 Photoshop 等软件对原始图片进行删改等,通常也只有依靠数字技术来识别和查证。因此,数据库服务商应参与开发或主动采用适当的软件工具,以遏制存在问题文献的传播。

5.2 开展与学术规范相关的研究

针对科研不端行为的技术检测,不可避免地涉及对有关文献的甄别和对确实存在问题的文献和责任人的处理工作。由于目前我国的学术规范和出版规范等尚不健全,在对有问题的论文和相关人员处理方面的政策也不完善,例如对存在抄袭剽窃问题的论文,大部分只是给予退稿或警告等处理,从而可能给人造成了“对学术抄袭剽窃问题法不责众”的印象。对于这些方面的问题,需要有关部门和单位在实践与研究中逐步摸索出适当的解决办法。数据库服务商由于能够接触到大量的案例和学术期刊、用户的信息反馈,因而有条件也有责任配合相关管理部门开展相应的研究,包括提出适当的政策建议。

5.3 保证学术文献的质量

维护学术文献的可靠性,是学术期刊单位和数据库服务商的共同责任。学术期刊应当将关于存在问题论文的更正启事和撤稿声明等及时通知数据库服务商,而数据库服务商对于利用技术

手段发现的存在问题的论文,也应向相关学术期刊或有关单位通报。目前,北京万方数据股份有限公司便已主动筛查或检测送交到中国期刊论文数据库和中国学位论文数据库的文献,并将结果提供给期刊社或有关政府部门、项目管理机构参考。考虑到错误的研究成果信息会对科研活动造成影响与危害,数据库服务商应当与学术期刊编辑部和其他学术出版机构共享抄袭文章列表,并通过互联网进行公布,以防这些文献被继续引用。

6 结 语

数字化时代为科技创新、知识传播提供了越来越丰富的手段和方法,学术数据库服务商高度重视和充分利用应对科研造假和抄袭剽窃等技术手段,将有利于促使科研人员和大学生严格遵守科研行为规范,有利于维护学术文献的真实性和可靠性,尤其是扩大技术手段在学术出版和科研项目立项验收中的作用,将提高我国科研水平和成果质量,进而促进我国科研竞争力的提高。

参考文献

- [1] 山崎茂明. Scientists Misconduct : Fabrication • Tampering • Plagiarism [M]. Translated by Yang Jian, Cheng Yuanyuan, Yan Lingna. Beijing: Tsinghua University Press, 2005. (in Chinese)
〔山崎茂明. 科学家的不端行为: 捏造·篡改·剽窃 [M]. 杨舰, 程远远, 严凌纳, 译. 北京: 清华大学出版社, 2005.〕
- [2] Li Zhenzhen. Chinese Science in Transition: Misconduct in Science and Analysis of the Causes [J]. Science Research Management, 2004, 25(3): 39-44. (in Chinese)
〔李真真. 转型中的中国科学: 科研不端行为及其诱因分析 [J]. 科研管理, 2004, 25(3): 39-44.〕
- [3] 综合 Turnitin 官方网站上“关于我们”、“首页”的相关介绍 [EB/OL]. [2010-09-18]. <http://www.turnitin.com/static/index.html>.
- [4] Hu Yuan, Jiao Lingzhi. Research of Free Open-source ROST System [EB/OL]. [2010-09-18]. http://www.sciencenet.cn/m/user_content.aspx?id=291380. (in Chinese)

(下转第 48 页)