

大数据时代网络文学版权保护的新问题及其规制*

王志刚¹⁾ 度冉²⁾

1) 中国海洋大学文学与新闻传播学院, 266100, 山东青岛; 2) 河南职业技术学院, 450000, 郑州

摘要 大数据技术是把双刃剑, 它在为网络文学产业带来发展机遇的同时, 也为网络文学版权保护带来新问题。无论是文学生产层面的智能化自动抄袭、无限制的同人文写作, 还是传播层面盗版网站对大数据分析结果的利用, 都将是新时期网络文学产业发展面临的难题。应尽快建立版权信息共享的统一检测平台预防侵权, 采用区块链技术提升维权, 加强大数据人才储备, 并在宏观层面提供有力的制度支持。

关键词 大数据; 网络文学; 版权保护

DOI:10.16510/j.cnki.kjycb.2018.11.006



开放科学 (资源服务)
标识码 (OSID)

2018年9月第二届中国“网络文学+”大会发布的《2017年中国网络文学发展报告》显示, 我国网络文学读者已突破4亿, 平均阅龄为4.9年^[1]。这些带有“数字原生代”特征的读者群体, 在阅读中也在为网络文学产业贡献数据积累, 使得大数据技术在网络文学读者细分、选题优化、衍生开发等方面有了更大施展空间。不仅如此, 大数

据似乎也为一直困扰网络文学产业发展的版权问题带来“福音”, 不断有消息报道一些企业利用大数据追踪抄袭盗版根源、用DNA技术维权等。然而, 我们要切记的是, 技术始终是柄双刃剑: 大数据可以为版权骑士所用去斩妖除魔, 也可以被版权盗贼拿去成为侵权利器。据最高法院发布的《知识产权侵权司法大数据专题报告》显示, 目前著作权侵权案件在所有知识产权案件中占比50.2%, 约6 000件, 其中3/4属于侵害作品的信息

* 基金项目: 本文为国家社科基金项目“大数据时代网络文学版权运营及保护研究 (17BXW101)”的阶段性成果。

作平台, 开辟更多合作渠道, 探索中国文化更有效的传播路径, 从而实现中文在线向海外市场的进一步扩张, 提升中国在全球范围内的文化软实力和国际话语权。另外, 中文在线将更加积极地推动更多创新文化产品持续落地海外市场, 搭建起中国文化与世界交流的桥梁。

参考文献

- [1] 李贞. 基于提高我国文化软实力的出版国际化战略研究[J]. 出版广角, 2015(13): 18-20.
- [2] 约翰·费斯克. 理解大众文化[M]. 北京: 中央编译出

版社, 2001.

- [3] 塞缪尔·亨廷顿. 文明的冲突与世界秩序的重建[M]. 北京: 新华出版社, 2010.
- [4] 习近平. 在中国共产党第十九次全国代表大会上的报告[EB/OL]. [2017-10-28]. <http://cpc.people.com.cn/n1/2017/1028/c64094-29613660.html>.
- [5] 李克强. 代表国务院在十三届全国人大一次会议上所作的《政府工作报告》[EB/OL]. [2018-03-22]. http://www.gov.cn/xinwen/2018-03/22/content_5276608.html.
- [6] 雷鹤. 中文在线“文化+教育+”发展战略启示[J]. 产业驱动, 2018(5): 20-24.
- [7] 刘蒙之, 贾瑞雪. 中文在线的数字出版品牌建构思考[J]. 出版广角, 2013(增刊1): 104-106.

(责任编辑: 彭远红)

网络传播权、放映权的案件。^[2]这其中网络文学案件分量不遑多让。因此,我们必须认识到,既然大数据会对网络文学发展产生整体影响,那么也一定会为整个产业的版权保护带来一些新问题。

1 快速发展的智能写作软件为抄袭提供“神器”

抄袭一直是困扰网络文学生产的难题,虽然大数据技术可以通过数据追踪等方式为维权者提供证据,但是在某种程度上,大数据也为抄袭提供了更多难以管控的空间。

在网络文学生产方面,智能化写作带来的抄袭将成为主要困扰。传统抄袭者可以用手打或者复制粘贴的方式拼凑出一篇文章,但抄袭对象有限,想完成一部“精致抄袭”作品难度仍然很高。而智能化抄袭,只需使用一个基于大数据技术支持的写作软件,就将抄袭者从“烦琐”的工作中解脱出来。这些所谓的智能写作软件通常由自动写作系统和素材库构成。自动写作系统强调写作“自动化”,根据作者输入的关键词,不仅自动生成人名、地名、服饰、人物性格等,甚至故事的情节结构、矛盾冲突、具体的场景都能自动呈现。后台软件库里的素材数量庞大,内容从世界名著到杂谈随笔,应有尽有。这种智能化处理由于不是集中抄袭某一部作品,而是隐藏在众多数据之中,很难被发现且难以界定。而这些原本帮助作者获得写作灵感的数据软件,现在成了抄袭神器,致使网络文学抄袭范围更广,其危害也更大。

随着大数据和机械学习技术的快速发展,人工智能写作能力也在逐渐提高,一些模仿经典作品而来的智能化作品颇具可读性,有的甚至很难区分是人类还是智能写作,为作品创作主体的

法律认定设下难题。同时,一些高端软件具备的“洗稿”功能,使得这些自动写作作品远离现有的抄袭认定标准,这些本质上“大规模小范围”的抄袭写作,在版权技术层面却很难取证和认定。因此,智能化写作将为网络文学抄袭治理带来诸多法律和技术难题。

2 同人文写作的规模化扩张成为侵权治理难题

从钱钟书先生的《围城》到鲁兆明的小说《围城之后》,从金庸先生作品到江南的《此间的少年》,当前网络文学写作中突出的同人文创作现象,显然会引发诸多版权问题。

同人这一概念最早出现在日本的漫画领域,随后成为一种网络文学创作形式,它利用原创作品中的角色、故事情节、背景设定等元素或者真实的人物形象进行再创作。^[3]大数据时代的同人文写作早已不再是针对某一部作品具体进行改写或者续写,而是开展了“无限流”的写作。写作者根据社会热点或者热门网文作品,选择多个题材并由多个故事拼凑成一部新作品。这种作品也成为当前网络文学平台的重要内容板块。如阅文集团旗下的晋江文学城平台,专门为衍生/轻小说设立板块。在2018年2月7日晋江言情长生殿的前20名小说中(图1),有12部网络小说标有[综](即综合多部他人作品而成),只有一部《猎人同人——无处不在的龙套生活》是单纯的以日本漫画家富坚义博的作品为主来写,但其内容存在大范围相似的情节、人物。其他11本网络小说都是一文多事,借鉴多本书的情节框架。对于这一类同人写作者来说,由于作品剧情复杂、故事多样,改编成影视游戏难度较高,写作的最重要考虑是如何吸引读者,如何写出刺激



图1 2018年2月7日晋江言情长生殿前20名

读者购买欲望的文章。因此，这些写作者特别注意读者的意见，甚至在写新故事前会主动提供几种故事构想，根据读者留言选择相应的故事进行写作。一些作者甚至会故意选择一些热门的作品进行二次创作，侵犯原创作者版权是很难回避的问题。

不仅如此，相较于正规网络文学平台的相对受控，盗版网站的同人化创作则完全失控。这些非法平台为了增加点击量，针对部分热门作品，会根据读者的“痛点”，专门组织写手创作“后传”“前传”“外传”等形式的同人作品，大量引用原作剧情来吸引读者。无疑，同人借鉴写作的规模化扩张显现出当前版权管控的相对无力，未来也会带来更多新问题。

3 大数据分析结果被盗版网站利用

相较于抄袭而言，盗版对网络文学伤害更大，而各大网文平台通过大数据分析并且不断更新的热文榜恰恰为盗版平台提供了盗取对象。

榜单中一目了然的数据，在方便读者选书的同时，也使盗文网站明确了受众的关注点，随即知道了机器抓取或者手打的重点所在。以晋江文学城签约作家时镜为例，他的《我不成仙》在修仙类作品中排名靠前且正在付费连载。然而在360搜索“我不成仙”时，首先出现的就是女生文学、晋江文学城、平板电脑三个文学网站（图2）。在

晋江文学城平台上，作者仍需花钱购买该书的VIP章节，然而女生文学、平板电脑的章节和晋江文学城不仅完全一致而且全部免费。运用百度搜索“我不成仙”，在搜索结果中虽然晋江文学城的链接位居首位，但随后是各种宣称“无弹窗广告”的盗版平台。当章节数一致、内容一致甚至盗版内容首先出现时，几乎不用考虑，大部分读者会选择阅读盗版书籍。

我不成仙_最新章节_全文阅读

作者：时镜
类型：古风古言
状态：连载中
简介：——纵你成仙，也逃不出我这一剑。她持剑，纵横捭阖，无数仙人在她剑下丧命。于是，好事者多为之曰：仙见愁...
最新章节：425 第425章 思辨与魔障 2018-01-19

女生文学	晋江文学城	平板电脑
425 第425章 思辨与魔障	424 第424章 弱肉强食	423 第423章 对质
422 第422章 妖性	421 第421章 生识不死离	420 第420章 大妖的逻辑
419 第419章 盛怒	418 第418章 烂柯楼论道	417 第417章 也过问心
416 第416章 荒古遗种	415 第415章 一剑胜有界	414 第414章 隐忧

图2 360搜索《我不成仙》显示结果

更为可悲的是，读者似乎只看重最终的免费阅读结果，丝毫不会考虑作品来源是否合适。艾瑞咨询《2016年中国网络文学行业研究报告》显示，27.1%的用户不清楚看的是正版还是盗版，44.7%的用户表示正版盗版都会看^[4]。当免费盗版章节的视觉效果和正版几乎没有差别的时候，似乎不难理解看盗版网文受众的心理，而盗版平台也乐意利用技术紧跟网络文学热点，满足读者旺盛的“免费阅读”需求。

4 大数据时代网络文学版权保护的建议

我们可以利用大数据分析对侵权行为进行追溯，找到发生盗版行为的关键所在。但是，当我们被海量数据包围时，如果硬件设施、技术手段以及其他配套条件没有达到相应的水平，就很难去分析所有的数据，找到正确的侵权主体。因此大数据时代网络文学的版权保护，需要平台、技

术、人才和制度等方面有机结合。

4.1 搭建版权信息共享的大数据监测平台

建立基于大数据技术并涵盖众多网络文学企业的版权信息共享平台，有助于管控智能软件写作和同人借鉴，也有助于监测盗版行为。

虽然一些网络文学平台很早开始搭建数据共享平台，但作为以盈利为目的的企业，不可能无私分享所有数据。即使有平台可以通过数据分析找到抄袭盗版源头，但由于不是官方执法人员，可能被侵权者无视，甚至还给部分非法传播网站经营者提供了毁灭证据逃避处罚的机会。

中国的剑网行动取得较好效果也正是依赖于国家级版权监测中心平台的应用，在这一平台上可以快速进行版权认证、预警、监测和维权。因此如果有专门针对网络文学的正版版权大数据中心和全网监测系统，将各文学网站的内容资源整合起来，可以瞬间发现上传的盗版内容，直接过滤或删除未经授权的网络小说。甚至还可以像论文检测一样，通过对比分析找到文章中所有的抄袭段落乃至句子，直接固定证据，避免了人工手动制作“调色盘”的麻烦。

4.2 利用区块链技术系统追踪维权

沉睡的数据一无是处，而能被充分利用的数据堪比黄金，只有通过技术手段对数据进行挖掘分析，才能得出有用结论。对于网络文学来说，不仅要加强大数据的基础设施建设，更要加强其应用程序建设，提升从大量用户数据中获取关键信息的能力，以便通过预警、实时感知、实时反馈等方式做出更好的战略决策。为此，我们要不断提升数据挖掘、数据分析、数据可视化等技术水平，利用技术保护措施打击侵权行为。北京奥运会新媒体版权保护就是利用技术遏制盗版的典

型，通过地域保护技术、过滤技术和追踪技术，它不仅限制指定地区以外的用户访问，过滤或删除未经授权的视频文件，还通过检测网络传输的视频信号追踪到侵权的IP地址，实现了“零盗版”这个几乎不可能完成的任务。

对于网络文学而言，由于很多作者写文时并没有版权登记的意识，很多作家都是等到作品需要进行IP改编时才会进行版权登记活动，一旦遭到侵权就必须面对“证明我是作者本人”的难题，而区块链技术给解决这一难题提供了一剂良药。区块链版权登记利用区块链去中心化的技术特征，为原创作品嵌入16进制的密码，这个密码会同时储存在区块链的所有电脑上，即相当于为原创作品登记了一张“电子身份证”，且永久有效，无法篡改。^[5]相比于传统版权登记方式高昂的登记费用和较长的受理期限，区块链版权登记不仅可以快速登记版权，在发生诉讼时，由于嵌入的独一无二的“密码”贯穿版权交易始终，使版权所有者可以快速获取固定侵权证据，利于自身维权。区块链技术早已被用到房屋产权证明、学生学历证书和成绩录入事宜中，不仅避免造假，还降低了成本。在中国版权保护中心DCI（基于数字版权唯一标识技术）体系中以及在贵阳大数据交易所里，我们也可以看到区块链技术的身影，在数据确权、数据溯源发面发挥了重要作用。对于网络文学而言，建立数字身份证，利用大数据技术提前过滤或删除盗版链接，可以实现事先预警、事后快速维权。

4.3 加强大数据版权保护人才培养

大数据技术对医疗、金融、人力资源、天气乃至预测犯罪等人类生活的全方面产生重要影响，数据分析专业人员也随之受到热捧。他们大多具有编程、计算机科学相关的专业背景和使用

SPSS、SAS等主流统计分析软件的技能, 现有的人才很难以满足市场需求。据全球顶尖管理咨询公司麦肯锡(McKinsey&Company)在2016年12月发布的分析报告显示: 预计2018年大数据或者数据工作者的岗位需求将激增, 其中大数据人才的缺口在14万~19万人, 而未来10年内, 仅美国对这类人才的需求量约为200万~400万!^[6]

利用大数据实现版权保护更需要人才支持。阅文集团之所以能够起诉多家盗版网站胜诉, 正是因为他们自己开发出了以文学作品数字指纹技术为核心的版权追踪系统。该系统不仅自发识别网友上传的作品是否侵权, 还可以纪录盗版信息, 为维权固定证据。中文在线投入了大量技术骨干并花费5年多时间研发出移动出版版权保护应用系统和网络侵权追踪平台, 在为企业维权方面立下赫赫战功。冠勇科技的版权大数据监测平台为超过500家权利人做版权监测服务, 130多位工作人员、500台核心服务器一天监测超过2万件影视作品、上百万首音乐作品和数百万篇文学作品, 每年光研发投入就超过5 000万元。^[7]而创造《纸牌屋》全球观剧热潮的美国视频网站Netflix, 曾两次设下百万美元奖金, 分别探求使推荐系统上升10%的推荐算法、架构和为不经常评级或者不评级的顾客推荐影片的方法。众多知名文化企业版权业务都有雄厚的大数据人才支持, 而中国网络文学企业作为国内IP产业链的重要发起点, 更是急需加大大数据人才培养投入。

4.4 完善大数据版权保护宏观制度

国家版权局在《版权工作“十三五”规划》中明确提出版权工作要坚持实施版权严格保护、推动版权产业发展、健全版权工作体系三大基本原则, 实现“加快版权强国建设, 为建成中国特色、世界水平的版权强国奠定坚实基础”这一战

略目标。^[8]而当前的版权制度环境, 虽然我国制订了《信息网络传播权保护条例》《关于加强网络文学作品版权管理的通知》等多部法律法规来明确网络服务商在版权管理方面的责任义务, 但针对大数据背景下的版权问题明显缺乏有效关注。

在制度层面, 我国目前对利用数据软件的写作行为以及对于同人作品“无限流”的写作方式等并没有明确的处罚措施, 甚至侵权标准都不明确, 客观上助长了部分侵权人尤其是不法平台的嚣张气焰。相较而言, 欧美国家针对大数据时代的版权新问题给予了快速反应。美国2010年“版权原则项目组”发表的《版权原则项目: 改革方向》提出将避风港规则中的“通知—删除”机制改为“通知—永久删除”机制等多项建议, 对大数据时代作品的版权获得、网络平台版权责任的加强等作出新的展望^[9]。这些建议在2013年正式进入众议院审议^①, 2016年已经完成全美听证^②, 目前部分议案已经进入立法程序^③。走的是更快的是欧盟, 在欧盟层面版权法《数字单一市场版权指令》中要求信息传播平台利用大数据技术严查内容抄袭和盗版传播, 提升网络服务商的侵权注意义务, 加大“避风港”规则适用难度^[10]。而我国原来征询意见的2013年版《著作权法》草案, 对于网络文学平台等传播者的技术责任要求过低, 因此在征询过程中引起强烈反对意见。而参照此时的欧美版权法改革, 我国《著作权法》更应该强调基于大数据的识别、追踪、取证等技术运用, 并将这一技术责任界定为网络服务平台适

① <https://judiciary.house.gov/pressrelease/chairmangoodlatteannouncescomprehensivereviewofcopyrightlaw/>.

② <https://www.copyright.gov/policy/>.

③ <https://judiciary.house.gov/press-release/goodlatte-conyers-release-first-policy-proposal-copyright-review/>.

用“避风港”规则的前提，以此强化平台版权义务，从外部制度要求去推动平台对智能化抄袭、无限制同人借鉴等新版权问题的主动监管。

5 结语

在版权制度演变过程中，每一个关键点都可发现技术革命的影子。15世纪金属活字印刷术的出现与普及催生了现代版权制度，19世纪后期全蒸汽动力的航海技术突破所带来的贸易全球化催生了国际版权保护公约，20世纪50—60年代的留声机、卫星等传播技术促进邻接权制度，新世纪前后数字网络技术的发展催生了一系列的数字版权法。而大数据以及在此基础上的人工智能技术的快速发展，也必然对保护创意和传播的版权管理提出更多制度要求。

从大数据技术诞生到其广泛被应用，我们所生存的这个数字社会已经被其所裹挟，所以，我们要做的，就是积极拥抱它。尤其是网络文学产业这个“数字原生”物种，从网络文学作品生产到网络文学作品传播再到网络文学受众分析，大数据技术可以将许多“不能”转变为“可能”。我们仍然不能绝对乐观，因为这种技术力量，对其积极利用可以成为产业发展的助推器，利用不当却可成为产业发展的绊脚石。对于大数据时代网络文学产业版权保护新问题的治理，既需要版权信息共享平台的统一监测以预防侵权，也需要不断纳入新技术提升维权实践能力，同时还需要培养大数据人才以提升新环境下的版权保护力量。更为重要的是，一个突出网络服务平台“侵权注意责任”、加大“避风港”规则适用难度的版权法律体系亟须建立，在这一方面，可以借鉴

欧美尤其是欧盟的相关版权立法措施。

参考文献

- [1] 路艳霞. 中国网络文学读者超四亿[N]. 北京日报, 2018-09-17(2).
- [2] 司法大数据研究院. 知识产权侵权司法大数据专题报告[EB/OL]. [2018-09-28]. http://www.court.gov.cn/upload/file/2017/10/13/16/35/20171013163521_97700.pdf.
- [3] Herzing M J. The Internet World of Fan Fiction [EB/OL]. [2018-09-28]. <https://core.ac.uk/download/pdf/51291784.pdf>.
- [4] 上海艾瑞市场咨询有限公司. 2016年中国网络文学行业研究报告[EB/OL]. [2018-09-28]. <http://report.iresearch.cn/report/201603/2540.shtml>.
- [5] 申屠晓明. 传媒行业区块链应用模式与技术方案解析[J]. 传媒评论, 2018(4): 27-31.
- [6] Henke N, Bughin J, Chui M, et al. The Age of Analytics: Competing in a Data-Driven World[EB/OL]. [2018-09-28]. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world>.
- [7] 李明远. 冠勇科技: 面向全球提供版权大数据服务[N]. 中国新闻出版广电报, 2017-08-10(6).
- [8] 中华人民共和国中央人民政府. 国家版权局印发《版权工作“十三五”规划》[EB/OL]. [2018-09-28]. http://www.gov.cn/xinwen/2017-02/16/content_5168508.htm.
- [9] Samuelson P. The Copyright Principles Project: Directions for Reform, 25 Berkeley Tech. L.J. 1175 (2010) [EB/OL]. [2018-09-28]. <http://scholarship.law.berkeley.edu/facpubs/563>.
- [10] European Commission. Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market (2016/0280 (COD)). at 29 (Sep. 14, 2016)[EB/OL]. [2018-09-28]. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016PC0593&from=EN>.

(责任编辑: 彭远红)