

单位代码	10475
学 号	104753150463
分 类 号	G210

# 河南大學

## 硕士学位论文

大数据时代网络文学的版权保护

学 科、专 业：新闻学  
研 究 方 向：编辑出版  
申请学位类别：文学硕士  
申 请 人：度 冉  
指 导 教 师：王志刚 副教授

二〇一八年六月



单位代码	10475
学 号	104753150463
分 类 号	G210

# 河南大學

## 硕士学位论文

大数据时代网络文学的版权保护

学 科 、 专 业 ： 新闻学

研 究 方 向 ： 编辑出版

申 请 学 位 类 别 ： 文学硕士

申 请 人 ： 度 冉

指 导 教 师 ： 王志刚 副教授

二〇一八年六月

# Copyright Protection of Internet Literature in the Era of Big Data

A Dissertation Submitted to  
the Graduate School of Henan University  
in Partial Fulfillment of the Requirements  
for the Degree of  
Master of Arts

By

*Duo Ran*

Supervisor: Prof. Wang Zhigang

June, 2018

## 关于学位论文独创声明和学术诚信承诺

本人向河南大学提出硕士学位申请。本人郑重声明：所呈交的学位论文是本人在导师的指导下独立完成的，对所研究的课题有新的见解。据我所知，除文中特别加以说明、标注和致谢的地方外，论文中不包括其他人已经发表或撰写过的研究成果，也不包括其他人为获得任何教育、科研机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

在此本人郑重承诺：所呈交的学位论文不存在舞弊作伪行为，文责自负。

学位申请人（学位论文作者）签名： 度冉

2018年6月15日

## 关于学位论文著作权使用授权书

本人经河南大学审核批准授予硕士学位。作为学位论文的作者，本人完全了解并同意河南大学有关保留、使用学位论文的要求，即河南大学有权向国家图书馆、科研信息机构、数据收集机构和本校图书馆等提供学位论文（纸质文本和电子文本）以供公众检索、查阅。本人授权河南大学出于宣扬、展览学校学术发展和进行学术交流等目的，可以采取影印、缩印、扫描和拷贝等复制手段保存、汇编学位论文（纸质文本和电子文本）。（涉及保密内容的学位论文在解密后适用本授权书）

学位获得者（学位论文作者）签名： 度冉

2018年6月15日

学位论文指导教师签名： 王长刚

2018年6月15日

## 摘 要

大数据对社会现有的结构、体制、文化和生活方式等方方面面造成了巨大的冲击和影响，它在某种程度上颠覆了企业传统的管理模式。而网络文学产业在经历了十几年的发展之后，数字化基因使其在发展过程中形成了大量数据库。数据库中优秀 IP 的使用、传播过程中所产生的数据，同样为这个产业发展提供了坚实的大数据应用基础。但是，在网络文学产业发展过程中，版权侵权的问题屡见不鲜。尤其是近年来，随着精品 IP 的不断涌现，版权侵权行为出现在影视改编、游戏改编等多个领域，如何利用数据进行网络文学版权保护成为我们目前的当务之急。文章分析了利用数据进行网络文学版权保护的优势所在，同时深入探讨了现阶段网络文学产业在版权保护方面的问题和不足，最后在探寻大数据时代网络文学版权保护路径方面提出一定的建议。

大数据在网络文学版权保护方面具备一定优势。经过十几年的发展，阅文集团、阿里文学等小说网站积累了大量用户和热门小说数据，通过数据挖掘、数据分析、数据可视化等技术手段，可以从多个角度找到并显示出版权容易产生的问题，为抑制盗版出现提供经验，版权威胁的超前预判也渐渐成为可能，通过数据预判，将盗版扼杀在萌芽里即将成为现实。对于已发生的网络侵权行为，利用数据追溯技术，我们不仅可以查到具体被侵权的书目、用户点击量等信息，还可以根据点击率粗略估算出已经造成的经济损失，固定数据证据，形成全面详细的检测报告，开展一站式维权。

与此同时，大数据技术的应用也催生了网络文学版权新问题，突出表现为文学生产、文学传播两个方面的侵权问题及维权面临的新困难。在文学生产阶段，利用“写作神器”等软件提取数据、进行写作，似乎已经成为部分作家的“常态”。由于抄袭书目繁多、范围极广，很难被人发现。同时，由于国内目前对同人作品可以借鉴的程度并没有明确的规定，无限流等大量借用热门作品数据的写作方式能否被判定为侵权还存在较大争议。而在文学传播阶段，大数据预测作用被盗版网站窃取，非法复制、盗版链接以及海外自发翻译侵权的针对性更强。为了抢夺受众点击率，赢得市场，针对热门作品的侵权，早已扩展到了影视、动漫、游戏等多个方面，形成了完整的盗版产业。由于译者翻译文章主要是靠读者打赏或者众筹获得收入，因而在未取得原创作者授权的情况下，私自翻译热门作品并免费上传网文译文的现象层出不穷；部分译者甚至为了抢夺受众点击率，

出现恶性竞争、彼此侵权的情况。不仅如此，随着互联网的发展，技术的进步，在维权阶段，版权所有者也面临着新的困难。侵权作品扩散迅速，且由于现有的技术设备很难存储并处理所有的数据信息，电子证据又容易被销毁，导致侵权主体认定模糊，原创作者维权困难、诉讼效果差。

对于大数据技术应用引发的一系列网络文学版权保护新问题，应该从技术支持、人才培养、制度构建等方面采取针对性的应对性措施。具体而言，可以由政府牵头搭建大数据监测平台，实现版权信息共享；提升大数据技术水平，以便更好的过滤侵权信息和追踪维权；加大教育投入力度，培养大数据版权保护人才；加强顶层设计，完善相应的法律法规。

文章在结语部分给出了研究的基本结论，并指出未来的研究可能。

**关键词：**大数据，网络文学，版权保护

## ABSTRACT

Big data makes a tremendous impact on many aspects of current social system and structure, culture and lifestyle; it subverts the traditional management system of companies. With the development of network literature for one decade, it results in the huge database caused by its digital feature. The excellent IP and the data created by propagation support the development of database application. However, the issue about copyright in this industry is common. Especially in recent years, with the appearance of excellent IP, movies and games adaptation suffer from copyright infringement. Because of it, the main task is how to protect the copyright of network literature based on big data. This paper analyzes the advantages of protecting copyright through big data and the issues about the copyright protection. Finally, it presents some suggestions about how to protect copyright in the age of big data.

Big Data has certain advantages regarding network literature copyright protection. After more than ten years of development, websites such as Yuewen Group and Alibaba Literature have accumulated a large number of users and hot novel data. Through data mining, data analysis, and data visualization technologies, it is possible to find and display issues about copyright from multiple perspectives, which can curb the emergence of copyright threats has gradually become possible. By predicting data, stifling piracy in the cradle will be done. For online infringements that have occurred, using data traceability techniques, we can not only find specific infringements of bibliographies, user traffic, etc. but also roughly estimate the economic losses, fix data evidence, create a detailed inspection report to start the one-stop rights protection based on clicks.

At the same time, the application of big data technology has also created new problems in online literature copyright, highlighting the issues of infringement in literary production and literary dissemination as well as the new difficulties in rights protection. At the stage of literary production, it seems to be "normal" for some writers to use "writing artifacts" and other applications to extract data to write. Due to a large number of books copied and their extremely wide range, it is difficult to be discovered. At the same time, because there is no clear rule about which domestic works can be used for reference, it is still controversial about whether unlimited writing and writing of popular works can be judged as infringement or not. At the stage of literary dissemination, the predictive ability of big data was used by pirated websites,

because of this reason, illegal copying, piracy links, and overseas spontaneous translation infringement were more targeted. To capture the market by grabbing the audience's click-through rate, the infringement on popular works has already been extended in many aspects such as film and television, animation, and games, which result in a complete pirated industry. Since translators of translate articles mainly obtain income rely on readers' rewards or crowdfunding, the phenomenon of privately translating popular works and uploading free translations on the internet without authorizing the original authors is endless; Some translators even have vicious competition and infringe each other in order to snatch the audience click-through rate. Besides, with the development of the Internet and the advancement of technology, copyright owners are also facing new difficulties during the period of rights protection. Infringement works spread rapidly, and existing technology and equipment are difficult to store and process all data information, electronic evidence is easily destroyed, which is leading to the inaccurate identification of the infringing subject, difficulty in defending the rights of original authors, and poor litigation effectiveness.

For the new issues of copyright protection of network literature caused by the application of big data technology, the targeted response measures should be taken from technical support, personnel training, and system construction. Specifically, the government can take the lead in setting up a big data monitoring platform to share copyright information; upgrading the level of big data technology to filter infringing activities and track the rights better; increasing education investment to cultivate talents for big data copyright protection; strengthening the top-level design, improving the corresponding laws and regulations.

In the conclusion of this paper, the basic conclusions of the study and future research point are presented.

**KEY WORDS:** Big data, Network literature, Copyright protection



# 目 录

摘 要.....	I
ABSTRACT.....	III
目 录.....	V
绪 论.....	1
(一) 研究背景和意义.....	1
1.研究背景.....	1
2.研究意义.....	2
(二) 文献综述.....	3
(三) 理论基础.....	5
1.网络文学.....	5
2.网络文学版权.....	6
3.大数据.....	9
4.大数据对网络文学的影响.....	11
(四) 研究方法.....	14
(五) 创新之处.....	15
1.研究视角新.....	15
2.观点新.....	15
一、大数据时代网络文学版权保护的优势.....	17
(一) 版权数据优势突出.....	17
(二) 版权威胁超前预判.....	18
(三) 全面的一站式维权.....	20
二、大数据时代网络文学版权保护的新问题.....	23
(一) 文学生产：内容抄袭与非法改编.....	23
1.内容抄袭范围扩大.....	23
2.同人作品侵权突出.....	26
(二) 文学传播：大数据预测作用被盗版网站窃取.....	28

1.传播方式：非法复制和盗版链接泛滥.....	28
2.传播范围：海外自发翻译侵权.....	31
(三) 维权面临新困难.....	33
1.侵权作品扩散迅速.....	33
2.侵权主体认定模糊.....	34
3.原创作者维权困难.....	35
三、大数据时代网络文学版权保护的建议.....	37
(一) 搭建大数据监测平台，实现版权信息共享.....	37
(二) 提升大数据技术水平，系统追踪维权.....	38
(三) 增大教育投入，培养大数据版权人才.....	39
(四) 加强顶层设计，完善相应的法律法规.....	40
结  语.....	41
参考文献.....	43
致  谢.....	47
攻读学位期间发表的学术论文目录.....	49

## 绪 论

### （一）研究背景和意义

#### 1. 研究背景

根据第 41 次《中国互联网络发展状况统计报告》显示,我国网络文学用户规模约为 3.78 亿,较去年底增加 4455 万,网民使用率为 48.9%;其中手机网络文学用户规模约为 3.44 亿,较去年底增加 3975 万,网民使用率为 45.6%。<sup>①</sup>网络文学的载体进一步向移动端转移,规模更大、覆盖群体更全、分布结构更合理的样本群体,为大数据的采集提供了更多的可能性。

2016 年是十三五规划的开局之年,2017 年则是实施十三五规划的重要一年,在某种意义上来说,这两年在中国网络文学版权保护及正版化进程中占有十分重要的地位。2016 年 1 月由北京海淀法院开庭审理的快播案,让大众第一次认识到了网络服务商的主体责任,避风港原则不再是包括网络文学盗版网站在内的第三方网络服务商脱罪的借口。同年底,明确了各网络文学服务商主体责任的《通知》正式下发,网络文学的版权环境进一步优化。

盗版,早已成为中国文化发展的毒瘤。其实,我国早已开始利用技术手段,打击网络侵权行为,维护网络版权秩序,净化版权环境。已经开展 13 年的剑网行动,针对网络侵权的热点,依托国家级版权监测中心平台,实施重点监管、分类规范,开展了各个领域的版权专项整治活动。仅“剑网 2017”专项行动,检查网站 6.3 万个,关闭侵权盗版网站 2554 个,删除侵权盗版链接 71 万条,收缴侵权盗版制品 276 万件,立案调查网络侵权盗版案件 543 件。<sup>②</sup>

大数据时代,随着数据技术的不断进步,中国网络文学版权保护取得了重大进展,查处了一批大案要案,净化了网络作品的版权环境。但是,随着强 IP 的出现,以网络

<sup>①</sup> 中国互联网络信息中心.第 41 次中国互联网络发展状况统计报告发布[R/OL].北京:中国互联网络信息中心,2018.01.  
[http://cnnic.cn/g9ywm/xwzx/rdxw/201801/t20180131\\_70188.htm](http://cnnic.cn/g9ywm/xwzx/rdxw/201801/t20180131_70188.htm).

<sup>②</sup> 金融界.“剑网 2017”专项行动关闭侵权盗版网站 2554 个  
[EB/OL].<http://finance.jrj.com.cn/2018/01/16155323954837.shtml>,2018-01-16.

文学版权为核心开发的作品，从创作阶段就已经开始遭遇侵权，更不用说影视化、动漫化、游戏化的作品了，侵权盗版主体更加多元、侵权手段更加隐蔽、侵权种类也更加多样，盗版网络文学利益链条朝着网格化发展……

2016 年全年，PC 端和移动端付费损失高达 79.8 亿元。<sup>①</sup>这严重挫伤了作者的写作热情，恶化了网络文学版权交易环境，网络文学版权保护面对着更加严峻的挑战。为此，通过数据分析、数据挖掘、数据可视化等技术手段将侵权问题呈现出来，对于加强网络文学版权保护力度，净化版权环境，有着重要意义。

作为业内领先的正版数字阅读平台及文学 IP 培育平台，阅文集团凭借着旗下海量的作家和作品资源，凭借其背后 6 亿用户大数据支持，通过对用户阅读偏好、阅读习惯、阅读时长、付费金额等数据进行汇总分析，挖掘出其中具有价值的信息，不仅优化了作品创作，为选题、开发、运营提供数据支持，实现个性化推荐；还可以实时监控数据，追溯盗版源头，为依法维权提供数字证据，保护自身合法权益。但是，由于当前大数据技术限制、不合理的人才结构，处于“信息孤岛”的版权数据以及相对滞后的法规制度，使利用大数据加强网络文学的版权保护还存在一定的难度，亟需克服。

## 2. 研究意义

### （1）理论意义

文化对于一个国家的发展至关重要。中国的网络文学经过十几年的发展，其影响力早已超出了文本的范畴，积累了众多国内外受众。加强版权保护，对于提升中国文化实力、增强文化自信有着重要的作用。但是，在大数据时代，如何充分发挥数据作用保护网络文学版权，几乎仍处于空白状态，缺乏相关理论支撑。利用数据加强版权保护，从技术层面出发，以一个新的研究视角关注网络文学侵权问题，对于完善网络文学版权保护的理论依据有着积极的意义。

### （2）现实意义

我们生活在一个几何级数据爆炸的时代，数据在生活中的方方面面发挥着重要作用，利用大数据技术保护网络文学版权有着重要作用。但是在当前社会，由于技术、成本、人才、政策等原因，数据与版权并没有想象中结合的那么紧密。本文在分析大数据对网络文学影响的基础上，介绍了利用数据进行网络文学版权保护的优势所在，同时深

<sup>①</sup> 艾瑞咨询.2016 年中国网络文学版权保护白皮书简版[R/OL].上海:上海艾瑞市场咨询有限公司, 2017.04.  
<http://report.iresearch.cn/report/201704/2971.shtml>.

入探讨了现阶段网络文学产业在版权保护方面的问题和不足，最后在探寻大数据时代网络文学版权保护路径方面提出一定的建议。

## （二）文献综述

知网上最早关于网络文学文章是 2000 年吴过的一篇《因特网中的文学天空——网络原创文学的回顾及展望(一)》，他从文学网站、网络写手等方面介绍网络文学的发展历程；但是，对于网络文学的起源，部分业界人士有着不同的看法。他们认为，网文的起源并不是发端于 1998 年痞子蔡在 BBS 上发表《第一次亲密接触》，而是在专属网文平台建立，并开始收费之后才产生的。宝剑锋，国内大型原创文学网站起点中文网的创始人之一，近日在知乎上发表《说网络文学已经有 20 年的，都是伪学者》一文。他认为网文最大的特性就是连载性，作家每日更新三四千字，粉丝长期追读评论，与传统文学相比，它在生产机制有着本质的不同。

严莉莉在《网络版权侵权问题探究》中，将网络著作权侵权行为归为三种类型，其中第三类侵权行为特指在网络上侵犯网络著作权的行为，而网络文学的盗版行为正是从第三种侵权方式着手。郑延培在《论网络文学作品的版权保护》中，认为网络文学侵权主要分为创作阶段、传播阶段、以及衍生作品的版权侵权，主要从网络服务商的角度提出对策；刘晓兰在《网络文学版权保护问题研究》提到，网络文学的侵权方式主要为盗版链接、恶性搜索和内容复制，可以从受众、技术、法律等角度对网络文学版权进行相关保护。但是，他们并没有重视数据的作用，忽略了数据作为基础性资源和战略性资源的价值。

张振宇和周莉的《“大数据出版”的理念、方法及发展路径》主要从“数字出版”和“大数据出版”的概念辨析入手，指出两者最大差异是后者能够通过对信息的数据化（结构化）处理，来实现出版资源的重新整合与最大化利用。他认为“大数据出版”给出版行业重塑自身的核心竞争力、资源整合与服务方式、版权保护与交易方式带来的新契机。但是这更多是从传统出版业的角度入手，通过数据发现传统纸质图书的新价值，并没有提及网络文学如何和数据结合。

大数据时代最重要的就是思维方式的转变，英国作家维克托·迈尔-舍恩伯格和肯尼思·库克耶在《大数据时代》一书中认为，大数据开启了一次重大的时代转型，它的核心就是预测，通过“样本=总体”的数据，挖掘并显示数据之间的相关关系，为做出



决策提供依据。其实，杜启杰在《大数据大战略》中，也提出大数据时代最重要的是思维的转变，不要再追求精确的数据，要重视混杂性，不要过于追求因果，要重视相关关系。

作为大数据关键技术之一，在《大数据挖掘》这本书中，作者谭磊用通俗易懂的语言向我们介绍了挖掘数据的基本算法，如何利用数据挖掘找出数据之间的规律，将价值变现。而大数据与网络文学结合，不仅有助于利用大数据技术保护网络文学的版权，还促进了版权的运营和开发。张金婵和孙丽的《大数据如何助推网络文学 IP 发展》介绍了大数据与 IP 开发的结合，从内容生产到产品链构建，借助数据的理性分析，实现精准定位，激发 IP 资源的最大价值。

但是数据并不是万能的。蔡莉和朱杨勇在《大数据质量》一文中提到，数据溯源追踪主要是通过起源链的记录信息来实现追踪目的，因而数据本身质量与结果紧密相关，制约着数据产业的发展。为此，一味的求大求全，忽视数据质量的做法是不可取的。张玉宏在《品味大数据中》不仅介绍了大数据的力量，更讲述了大数据的陷阱，提醒我们不要误把相关当作因果，也不要过度迷信大数据，要客观认识数据，批判性的接受大数据思维。段云峰在《大数据和大分析》一书中提出，我国数据挖掘多是基本算法，很少使用改良类的算法，且渗透的领域很少，分析方法需要有“质”的变革，要转变分析理念，让更多专家参与进来。

李婷和刘广登在《大数据时代数字版权保护体系的构建与完善》中，针对大数据环境下数字出版产业发展的基本现状，分析了目前数字出版版权保护所面临的难点和困境，在此基础上，从立法层面、技术层面、环境层面和意识层面入手，提出完善我国数字出版版权保护体系的对策。许波也的《大数据时代数字出版的困境与对策研究》一文，从完善政策体系，提高文化产业地位等九个方面进行了深入研究并提出具有可行性的建议。邵琰在《数字出版著作权保护策略研究》中，认为目前版权保护立法理念相对滞后，提出建立健全数字出版产业法，使数字出版物登记制度更加健全，这样可以也使数字出版著作权保护的水平和层次得到有效提升。

综上所述，从大数据的角度对我国网络文学版权保护探析论文并不多见，现有的文献大多是以网络文学版权、数字版权、大数据和传统出版相结合的角度为切入点对其进行研究，在一定程度上我国网络文学发展可以借鉴其相关的研究成果，但是必须在该基础上进行创新，不能简单的拿来主义。

### （三）理论基础

目前中国网络文学行业正处于移动互联网时代的发展初期,以网络文学 IP 为核心的生产,已成为行业营收增长的核心,但是盗版始终制约着网络文学的发展。在大数据时代,数据对于网络文学版权的发展起着越来越重要的作用,通过深挖、分析、汇总数据得出的信息,不仅有利于版权运营管理,还可以固定数字证据,保护作者和版权所有者的利益,促进网络文学更好的发展。

#### 1. 网络文学

##### （1）网络文学的定义

网络文学是指以计算机和互联网为媒介而存在和传播的文学。<sup>①</sup>也就是说,只要在网络发表的文章都可以称为网络文学。然而,不仅学界和业界未达成一致,即使在学界中间,也存在着不同的看法。他们矛盾的焦点主要集中在两个方面:一是已经通过纸质出版的文章,以电子扫描或者人工输入的方式,发表在网上,能否称之为网络文学;二是凡是首次发表在网络上文章是否一定都是网络文学。

大多数学者认为网络文学发端于 1998 年,在这一年,痞子蔡在 BBS 上发表《第一次亲密接触》。但是与其说这是一种网络文学,不如说它是传统文学在网络上的变形,是传统出版物更文方式的一种变式。痞子蔡只不过是將线下文学的更文方式转到了线上,线上无法取得收益的作者,还是將重点放在纸质出版上,更关注纸质出版的收益。加上当时互联网并未普及,人们还习惯于传统纸质的阅读方式,网络上被转载的文章并未影响作家纸质图书的销售,因而很少有作家会因为自己文章在网上被人转载而提出异议,他们用着以前的模式对待着正在萌芽的网络文学,仍然保持着作者本位的思想。

在笔者看来,网络文学自然带有网络互动性的特征,本质上,它的产生过程应该是网站与作者、网站与读者、作者与读者、读者与读者之间交互影响的过程,是主体多元参与、互相交流、相辅相成、共同创造的一种过程。而这一时期发表在网络上的文章,恰恰缺乏这一点,即使刊登载体发生了改变,其本质仍然属于传统文学的范畴,并未发生根本性变化。

网络文学既不是传统图书出版之后经过电子扫描上传到网站上的文章,也不是在网络上发表的原创文章随笔,它是作者在专有小说网站上按期更新,读者付费阅读浏览相

<sup>①</sup> 欧阳友权.网络文学本体论[M].北京:中国文联出版社,2004:1.

应章节的小说，是在文学产业链形成之后产生的。这既不是等待作者完全写完一本书才能浏览文章全貌的更文方式，也不是作者凭借心情在网络上随写随发的更文方式，它带有一定的商业性和紧迫感，为了吸引受众，避免流失读者，作家不得不坚持更文；与传统文学的写作方式相比，尤其是进入大数据时代以后，网络文学发生了很大的变化。作者本位被打破，从此进入了读者为王的时代。如果作家完全不顾读者喜欢什么，不去了解自己作品更适合在那个网站上连载，除非文笔情节特别好，故事内涵特别深刻，故事案例特别新颖，否则单单凭借个人的兴趣爱好，很难激起读者的阅读欲望，也很难写出符合大众审美的热门作品。

## （2）网络文学的发展阶段

根据网络文学不同发展阶段特征，特别是根据用户阅读方式的不同，艾瑞咨询将网络文学分为 PC 时代、过渡时代和移动互联网时代三个发展阶段。<sup>①</sup>跟艾瑞咨询不同的是，笔者认为《第一次亲密接触》并不是网络文学的开端，真正的网络文学是在作家每天坚持进行更新，粉丝长期追读并进行评论，在作者与读者开展频繁互动的基础上产生的。当作家在网络上不再单纯的以个人的兴趣、爱好、话题为主发表文章的时候，当一个以网络文学发布为主营业务的网站、以虚拟书籍书库为阅读模式的书城模型正式建立的时候，当 2003 年起点中文网正式将作者和读者、内容与讨论进行分割的时候，当网络文学的版权走上专业化运营渠道的时候，“真正”的网络文学就产生了。

随着科技的发展，手机逐渐成为人们生活中不可或缺的一部分，当在网络文学手机端的阅读量首次超过 PC 端的阅读量时，过渡时代就正式结束了，我们从此进入了移动互联网时代。仅 2017 年上半年，移动端网络文学的月独立设备数均超过 2.2 亿，平均每月的有效使用时长超过 15 亿小时。<sup>②</sup>经过十几年的沉淀，网络文学集聚了大量的优秀 IP，以及庞大的粉丝群体，电影《致青春》，电视《花千骨》，动画《择天记》，游戏《甄嬛传》等由网络文学改编的作品都取得了不俗的成绩，强大的粉丝号召力在这一年得到巨大的体现。作为 IP 开发的源头，加强对优质网络文学版权保护力度至关重要。

## 2. 网络文学版权

### （1）网络文学版权的定义

<sup>①</sup> 艾瑞咨询.2016 年中国网络文学版权保护白皮书简版[R/OL].上海:上海艾瑞市场咨询有限公司,2017.04.  
<http://report.iresearch.cn/report/201704/2971.shtml>.

<sup>②</sup> 艾瑞咨询.2016 年中国网络文学版权保护白皮书简版[R/OL].上海:上海艾瑞市场咨询有限公司,2017.04.  
<http://report.iresearch.cn/report/201704/2971.shtml>.

网络版权是指版权所有者对其创作的受版权法保护的文学、艺术和科技作品在网络环境下所享有版权权利。以此类推，我们可以给网络文学版权下一个定义，即作家对其创作受版权保护的文学作品在网络环境下所享有的版权权利。

## (2) 网络文学版权的起源

在上文我们已经对网络文学进行了界定，认为真正的网络文学是在起点中文网开始根据章节的质量和字数以 2 分/千字的稿费制度支付给原创作者之后才正式出现，当文学产业链初步形成之后，中国才有了真正意义上的网络文学版权，从此，在网络上任意转载文章的现象将不复存在。经济收益使作者和原创网站开始重视网络文学版权的价值，不仅作家本人可以从中获得收益，网站为了吸引用户的点击量，获取格外的广告收益，也不允许其它网站转载自己原创作家的作品，网络版权正式进入人们的视野。

早在 1997 年，美籍华人朱威廉就创办了最早的中文原创网络文学网站“榕树下”，随后，黄金书屋、多来米中文网等文学网站也紧跟着出现。但是，在这一时期，无论是在 BBS、还是在集创作、更新、BBS 于一体的文学网站上，所有文章都是免费阅读的，BBS 和各大网站之间，文章也是可以任意转载，并不需要给作者、BBS 或者其它网站任何版权费用；对于读者来说，他们几乎不用花费任何力气，就能在网络上找到自己所有喜欢的文章章节，他们所花费的仅仅只是网费、电费，最多再加一个电子设备的费用；对于这一时期的作者来说，他们不在乎网络上文章被转载，他们更看重的是实体出版，网络只是他们写作的平台，借助这种新的载体来进行写文。BBS、文学网站等的转载甚至提高了作者自身以及文学作品的知名度，使他们的作品更易于出版和售卖。在这一时期，香港的出版社被无数来自内地的投稿所“掩埋”，但实际上，能够转变为纸质读物的毕竟是少数。没有经济利益，就没有纠纷。只有实体出版才会能给作家带来收益，为此作家看重的实体版权，网络上的版权基本上不存在。

其实盛大文学并不是收费模式的开创者，博库、星骸等原创文学网在它之前就已经开始收费，但是因为受众没有办法接受收费的形式，后续无力，难以维持下去。而起点文学网站开始以固定价格向用户收取阅读费用时，也曾受到用户抵制，但是凭借着一批优秀作家和优秀作品的加入，凭借着优质的 IP 资源，吸引了大量受众，顽强的生存了下来。收费模式开启之后，各大网络文学网站纷纷开始借鉴，网络文学的阅读收费从此成为国内各大主流文学网站的主要盈利模式之一。因此，我们可以说起点文学网开创了用户付费网络文学商业模式。

原创作者有了自己固定的写文平台，平台聚集了大量的受众，客观上扩大了作家的影响力。与此同时，作家本人也可以从中获得一定的稿酬，随着小说阅读平台作家福利制度的不断完善，原本可能是凭借一时兴趣写作的兼职作家，获得了经济上的动力，这大大提高了他们的创作激情。当沧月在 blog 更新《羽》系列时，他的更新周期大概维持在 4-7 天/次。<sup>①</sup>当网文进入付费阶段时，为了“赚钱养家”的唐家三少创下了连续 96 个月未断更的纪录。当网络作家可以凭借写文养家的时候，已无后顾之忧的他们可以尽情沉浸在网络文学的世界中，进行高质量的创作，写出优秀的作品。

### (3) 网络文学盗版

网络文学盗版是指在未经网络文学作者或版权所有者、版权收益方、相关运营站点等同意或授权的情况下，出于经济等目的，私自转载、发布、盗取或出版其相关作品的违法行为。<sup>②</sup>随着移动互联网时代的到来，在方便作者写文，读者阅读文章的同时，盗版技术也逐渐更新升级，跟传统的方式相比，不仅减少了纸张、机器等方面的生产成本，也减少了宣传、发行等方面的运营成本。人工打字、机器抓取，盗版方式更加多样，技术手段不断升级，秒盗似乎也成为网文界的常态。为了获取流量，盗版网站甚至会提前放上空白章节，出现盗版网站比原创网站“早更”的现象。部分作家由于心血被大量复制转载，愤而放弃写文，甚至有的直接放言，等待盗版网站更文。盗版严重伤害作者的写作热情，最终影响的还是读者利益，从长远看不利于网络小说的良性发展。

同时，随着中国网络文学不断向海外进行扩张，Wuxiaworld、Gravity Tales、起点国际等网文翻译网站不断涌现，培养了大批喜欢中国网文的外国读者，甚至出现了一位因为看网文而戒掉毒瘾的外国少年。实际上，在这些翻译网站上，很多中国网络文学的译作并没有取得原创作者的授权，译者出于分享的目的，自发翻译网文上传，导致国外网友习惯了阅读免费的翻译小说，一旦开始收费可能会失去大批受众。因为盗版网络文学客观上扩大作品影响力，增强了作者本人的知名度，加之地域文化语言的差异使跨国维权艰难，因此很多作家选择了沉默。

其实，现今我们所说的网络文学盗版绝不仅仅只是对于文字作品本身的盗版，随着《三生三世十里桃花》《楚乔传》《择天记》《全职高手》等影视、动画等衍生作品的大卖，围绕着原创 IP 相关的盗版活动层出不穷，盗版的链接、视频、游戏不断涌现，

<sup>①</sup> 社微世 Service.我眼中的 IP 发展史:网文作品的兴起史 [EB/OL]. [http://www.sohu.com/a/204094027\\_100049702](http://www.sohu.com/a/204094027_100049702),2017-11-13.

<sup>②</sup> 艾瑞咨询.2015 年中国网络文学版权保护白皮书简版[R/OL].上海:上海艾瑞市场咨询有限公司,2016.01. <http://report.iresearch.cn/report/201601/2515.shtml>.



如图 1 所示。



图 1

根据艾瑞咨询的估算数据，2015 年全年，中国盗版网络文学，如果全部按照正版计价，仅 PC 端和移动端付费阅读收入合计达到 79.7 亿元；2016 年全年，PC 端和移动端付费阅读收入更是高达 79.8 亿元，而这仅仅是线上对文字侵权所导致的金钱损失。网络文学与影视、动漫、游戏等多方衍生联动的基础就在于版权，版权已成为泛娱乐化的核心竞争力，其所面临的侵权问题不容忽视。作为开启了重大时代转型的大数据，无疑为解决网络文学盗版提供了一个新的思路。

### 3. 大数据

#### (1) 大数据的定义

数据是指使用约定俗成的关键字，对客观事物的数量、质量、位置及其相互关系进行抽象表示，以适合在这个领域中用人工或自然方式进行保存。<sup>①</sup>狭义的大数据是指用现有的一般技术难以管理的大量数据的集合。研究机构 Cartner 认为，大数据是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。<sup>②</sup>

与以前仅强调数据的“质之高”不同，它更重视数据的“量之大”，从 b、B、KB、MB、GB 发展到 PB、EB、ZB 甚至 YB，如今每天产生数据量远远超过了我们的想象。强调“量”，这并不是说精确的数据不再重要，而是因为在大数据时代，如果过早的划定一个标准，一开始就把很多可能性排除在外，导致结果出现偏差。没有人知道为什么把啤

<sup>①</sup> 郎为民.漫话大数据[M].北京:人民邮电出版社,2014:53.

<sup>②</sup> 周苏,王文.大数据导论[M].北京:清华大学出版社,2016:6.

酒和尿不湿放在一起会提高收益,但是如果一开始认为尿不湿和啤酒毫无关联将其筛选出去,我们可能就不会发现其中的关联。

大数据的核心是预测,相对于为什么,它更加重视是什么。当我们拥有大量数据并有着明确目标的时候,即使在分析过程中,存在着部分错误数据,影响也并不大,因为当数据大到一定程度之后,数据会自动分析并过滤掉这些错误的数据或者通过系统性建模消弱这些数据的影响,使我们无限趋近于正确的答案,达成预期的目标。因此,当我们掌握足够多数据的时候,错误的数据对我们的分析结果造成的影响微乎其微,即在大数据时代,样本本身就等于总体。

## (2) 大数据的特征

互联网的快速发展,社交媒体的广泛流行,使生活在 21 世纪的我们,处于一个被几何级爆炸的数据包围的时代。当然,这些跟数据存储和处理成本的降低也有很大关系。早在 2001 年的时候,高德纳分析员道格·莱尼就认为未来的数据增长将面临着量 (Volume)、速 (Velocity)、类 (Variety) 的挑战;<sup>①</sup>随后,着眼于数据应用的专家们在此基础上提出了大数据具有 4V 的特征,也就是现在我们所说的数量 (Volume)、多样性 (Variety)、速度 (Velocity)、价值 (Value),而“量”成为了大数据跟传统社会相比最大的不同点。

利用数据技术,短短几周天文学家收集到的数据已超过历史上的数据之和。在 2008 年的时候,我国手机网民规模约为 1.18 亿人,而到了 2017 年的时候,我国网民规模接近 7.53 亿,<sup>②</sup>10 年间人数几乎翻了 6.4 倍,以平均每年约 0.71 亿的人数增长,而这仅仅是手机移动端的数量,还不包括 PC 端等设备的。庞大的用户规模为我们获得足够多的数据提供了可能。而随着各种数据存储技术的升级换代,我们每天上网所浏览的文章、图片、视频等痕迹可以被电脑一一记录下来,当它们汇总在一起,大数据就产生了。

## (3) 大数据的作用

在大数据时代,数据是会给企业带来利润的资源,是所有一切商业行为开展的基础。2017 年双十一狂欢节刚刚开场 5 分钟,天猫达成的交易为 1536 万笔/分,而数据库处理峰值更是高达 252000 万次/分,更不用说,双十一全天所达成的数据,以及数据库所需要处理的数据总量。这些结构化或者非结构化的数据,如果只是单纯留在数据库中,

<sup>①</sup> FineBI.再谈大数据的特征,感受数据之大[EB/OL].[http://blog.sina.com.cn/s/blog\\_ecfc26de0102vh3l.html](http://blog.sina.com.cn/s/blog_ecfc26de0102vh3l.html),2015-03-14.

<sup>②</sup> 中国互联网络信息中心.第 41 次《中国互联网络发展状况统计报告》发布[R/OL].北京:中国互联网络信息中心,2018.01.  
[http://cnnic.cn/gywm/xwzx/rdxw/201801/t20180131\\_70188.htm](http://cnnic.cn/gywm/xwzx/rdxw/201801/t20180131_70188.htm).

只是增加政府、企业等存储的负担,只有借助一定的技术手段,从复杂纷繁的数据中提取到有用的信息,找到数据彼此之间的关联,才能为我们所用,才能将沉睡的数据变现。

作为大数据关键技术之一,大数据挖掘是数据价值变现的重要工具。它是指有组织有目的地收集数据,通过分析数据使之成为信息,从而在大量数据中寻找潜在的规律以形成规则或者知识的技术。<sup>①</sup>与传统的数据分析不同,它所挖掘的对象不仅仅包括我们所说的文本,也包含了我们所说的音频、视频以及图片数据等等,即使是不规范的数据,也可以通过数据挖掘展现出来。它应用了数理统计、数据库和人工智能中的大量技术,除了自身算法渐趋成熟外,种类越来越多的可视化工具,以及多样化的应用需求,客观上都促进了数据挖掘技术的发展,也加速了数据变现的进程。

从数据中我们可以就看到客户的信息,看到产品的信息,了解如何对消费者进行有针对性营销。它就像是一座座巨大的金矿,等待着人们去探索去挖掘。淘宝首页上所显示的商品,可能正是你最近浏览过的,网易云音乐每日推荐里的音乐可能非常符合你的品味,今日头条更是根据你的口味推荐新闻……除了营销和金融领域,制造业、医疗服务业、教育行业以及各种政府机关等都出现数据挖掘的身影,数据挖掘甚至还帮助了奥巴马在 2012 年总统选举中取得了胜利。根据中国信息通信研究所统计,2014 年中国大数据市场规模约为 84 亿元,预计 2015 年中国大数据市场规模将达到 115.9 亿元,增速达 38%,2016 至 2018 年中国大数据市场规模还将维持 40%左右的高速增长。<sup>②</sup>

#### 4. 大数据对网络文学的影响

无论是占有大量的数据,还是追求成熟的算法,其最终的目的都是为了实现数据的商业价值,对网络文学来说,如何利用数据挖掘技术,将 15 年来文学网站积累的大量数据变现,使得人们的行为和情绪的细节化测量成为可能,是我们目前所要面对的重要问题。

数据挖掘技术包括分类、聚类、关联分析、链接挖掘、装袋和增强、序列模式挖掘、机器人学习、聚合挖掘、粗糙集以及图挖掘算法等多种算法。对于网络文学而言,使用最频繁的就是链接挖掘算法,即处理和分析互联网上的超级链接以及相关的网页信息。通过对网站的日志分析,我们可以清楚的知道网站用户在什么时间、什么地点、用什么操作系统和什么浏览器访问了网站的哪个页面,是否访问成功,以及离开网站去往哪里。

<sup>①</sup> 谭磊.大数据挖掘[M].北京:电子工业出版社,2013:23.

<sup>②</sup> 艾瑞咨询.2016年中国数据驱动型互联网企业大数据产品研究报告[R/OL].上海:上海艾瑞市场咨询有限公司,2016.12.  
<http://report.iresearch.cn/report/201612/2694.shtml>.

通过提升内容质量和内容触达效率帮助经营者重新评估自己的作品，通过用户洞察和精准投放指导作品优化和实现个性化推荐，这对网络文学的发展有着重要意义。

### (1) 优化作品和服务

作为一种数据资产，数据的价值可能远远超过了固定资产的价值，其最大的优势在于它能利用已知的数据推测出未来可能出现的情景，从而优化产品和服务，更好地为用户服务。

每一个网络文学网站在经过长时间的积淀之后，都会形成相应的风格，有着鲜明的特色，吸引着相类似的读者；而读者的汇聚使编辑倾向于签约这种文风的作家作品，促使更多作家写出类似的文章，如此循环往复。以阅文集团旗下的三家网络文学网站为例，红袖添香网注重女生言情类的文章，起点中文网注重男主升级流的文章，涉及纯爱、无CP 晋江文学城，题材更加多样化。如果不了解网站风格，盲目投稿，只会以惨败收场。若把晋江文学城大卖的纯爱类文章拿到红袖添香小说网站上主推，大部分受众肯定难以接受。对于作家来说，选择适合自己的写文平台，选择适合自己文风的受众，才更有可能发挥自己的写文优势。

表 1 红袖添香网站 4 月 14 号热销榜排行榜数据

序号	书名	作家	字数 (万)	总收藏 (万)	总点击 (万)
1	宠妻 108 式：韩少，狠狠爱	恍若晨曦	96.39	1.15	360.01
2	99 次心动，情迷首席纪先生	秦若虚	187.03	1.75	236.9
3	皇上，请您雨露均沾	miss_苏	395.97	9.2	184.24
4	逼婚 36 计，总裁的旧爱新妻	浮华尽褪	154.42	2.16	8.71
5	娇妻高高在上	唐渐浓	164.47	1.31	328.02
6	限时蜜爱，总裁强行撩妻 100 天	苏木颜	64.56	0.61	93.81
7	帝姬传奇：华都幽梦	叶落葵	47.31	0.42	56.62
8	蜜妻有点甜：吻安，总统先生！	腊笔小酱	66.07	0.73	187.39
9	专制隐婚，霸道前夫难招架	蔚然语风	81.51	0.58	115.3
10	首席心尖宠：早安，小甜妻	夏夜梦戈	62.58	0.57	99.83

以红袖天香小说网为例,笔者点开其 2018 年 4 月 14 号热销榜排名前十的文章,如表 1 所示,写文字数、总收藏量、总点击率等数据信息一目了然,而这仅仅只是红袖添香网站上部分书目最基本的信息,在后台,我们可以看到更多读者的具体浏览痕迹。当小说阅读网站数据积累到一定程度之后,当大量结构化或者非结构化的数据汇集在一起时,有用的信息就出现了,哪里是读者兴趣点一目了然。网站的编辑会明确知道自己的受众喜欢看什么类型的作品,自己的网站需要什么样的作家,什么样的推荐页面更适合读者的浏览习惯。对于热门的、有较大发展潜力的作品,网站编辑会主动进行分类推广,从点击、收藏、推荐、月票等数据上的硬性分类推广到本周强推、编辑推荐等软性分类推广,覆盖全站作品,保证优秀作品能够得到展示。

对于读者来说,其最看重的还是文章题材是否新颖,剧情发展是否符合逻辑,埋伏笔是否合理……想吸引读者眼球,就需要了解读者的兴趣所在。当多个读者反复阅读同一章节或者在文章某一部分停留时间较长时,这可能就是文章中最触动读者的地方。如果将这些数据汇总、分类,也许就可以得出有用的信息,指导作者优化作品内容,寻找新的创作方向。

相对于传统文学,网络文学受到读者的影响更大,读者和作者之间沟通交流更为频繁,甚至很多作者在开启新的文章时,会主动征求读者意见寻求下一部作品或者新章节的灵感,在写文过程中也会参考读者意见修改剧情。这其实与韩国电视剧边拍边播的制作方式有点类似,编剧会根据电视剧收视率的情况,修改剧本,以适应市场需要。网络文学其实也可以根据读者反应,修改剧情或者反套路剧情,及时发现自己文章的不足,修改不合理的剧情,使其高潮不断,牢牢抓住读者眼睛。

多数网文作家还是从新手小白慢慢起步,读者的评论、订阅、评分是对他们作品最直接的反应,但是过于散乱的评价很难让作者抓住重点。如果可以对这些数据进行深入地挖掘分析,不仅编辑很容易看出作品的市场前景,作家也能明白读者的期盼,知道怎样修改文章来增加文章热度、话题度和点击率。虽然这其中可能存在一些无效的数据,然而当数据不断更新且不断增多时,这些被充分利用的数据能够有效的帮助作者找到更精确的创作方向,提升精品 IP 的数量和质量。

## (2) 个性化推送

根据读者的生命周期,我们可以将其分为潜在读者、首次购买的读者、活跃读者、短期或者长期流失的读者。如何让潜在读者成功购买 VIP 章节,如何增加小说阅读网站



线上和线下的活跃度,避免其由短期潜水变为长期潜水甚至流失到其它的平台,根据读者浏览的数据信息,有针对性的推送文章,激发读者的阅读兴趣,增加其活跃度,增大其线上付费阅读的欲望十分必要。

读者在小说网站进行首次登陆的时候,一般网站都会要求用户填写性别、出生日期、省份和城市,甚至填写常用的微信号或者微博昵称。这并不是无用功,其实根据注册的性别,基本就可以判断用户大概喜欢书籍的种类,当然,这个结论并不严谨,它还需要具体的阅读行为等数据进行验证。用户浏览的图书种类,每个章节具体阅读时间、停留的时间、充值的频率、金额、付费次数等,通过相关算法基本上已经形成了粗略的“人物画像”,用户的文学喜好和阅读模式清晰可见。这些隐含着数据的信息反馈给作者,可以帮助其改进书籍的内容和结构;反馈给网站,网站运营者就十分清楚读者喜欢哪个类型的小说,他们会为哪些章节付费,由此为其推送相似的文章信息,比如说“看了该作品的读者还看了什么”,开展定制化服务。

腾讯将用户数据按照基本属性、社交兴趣、娱乐兴趣、媒体兴趣、浏览习惯、购买倾向等不同纬度细分为 3000 多个标签,能够组合定制 TA 画像。<sup>①</sup>通过这些数据信息,腾讯能够推算出用户的消费喜好,浏览兴趣等,在网站上寻找相似的目标人群,有针对性的推送广告信息,为用户提供定制化服务。虽然目前还处于待开发中,晋江网也已经提出通过为符合要求的签约作者提供数据分析,让作者更了解自己的读者,更了解网站读者整体特性。

大数据是一种资源,也是一种工具,它告知你是什么,而不告诉你为什么,简单又直接。这些数据不仅可以洞察消费者行为、实时监测传播效果,还可以推动网站及时调整投放模式,精准投放产生经济利润。随着数据开放程度逐步提高,网络小说 IP 价值不断被开发,利用数据促进产品服务优化升级势在必行,数据资产将进一步向商业化扩展。

#### (四) 研究方法

文章主要采用了文献分析法、案例分析法、归纳分析法三种方式。先在图书馆借阅了大数据的相关书籍,对大数据有了初步的认识。然后针对典型的侵权案例,结合网络上大量的文献资料,分析归纳得出大数据时代网络文学版权可遇到的问题,并提出相关的建议。

<sup>①</sup> 腾讯智慧.慕思走近腾讯智慧,碰撞家居品牌营销新火花  
[EB/OL].[https://mp.weixin.qq.com/s?\\_\\_biz=MjM5ODY2ODMyMg%3D%3D&idx=1&mid=2650366124&sn=31e61b05695f81fc5006be6ee7d9010b,2016-10-31](https://mp.weixin.qq.com/s?__biz=MjM5ODY2ODMyMg%3D%3D&idx=1&mid=2650366124&sn=31e61b05695f81fc5006be6ee7d9010b,2016-10-31).

## （五）创新之处

### 1. 研究视角新

现有关于网络文学版权的论文，多是从加强受众版权意识、完善法律法规、加强行业自律等角度提出相关意见建议，技术类的建议并不多。即使有些关于版权保护的文章，涉及利用大数据技术的，要么是从数字版权的角度，讨论如何保护数字作品的版权；要么是将大数据与传统出版业相结合，重点在于如何运营版权，发掘纸质出版物的新价值。本文在分析大数据对网络文学影响的基础上，介绍了利用数据进行网络文学版权保护的优势所在，同时深入探讨了现阶段网络文学产业在版权保护方面的问题和不足，最后在探寻大数据时代网络文学版权保护路径方面提出一定的建议。

### 2. 观点新

本文对网络文学和网络文学版权的概念进行了重新界定，从网络文学产业的角度出发，认为“真正”的网络文学和网络文学版权是在专属的网文平台建立并开始向作家支付稿费之后才产生的，也就是说，在 2003 年起点中文网成立之后，才有了网络文学；当起点中文网开始根据章节的质量和字数以 2 分/千字的稿费制度支付给原创作者之后，网络文学版权才正式产生。



## 一、大数据时代网络文学版权保护的优势

近年来,网络文学的发展可谓是如日中天,从作品最初的文字形态到如今的动漫、影视、游戏、有声读书等多种作品遍地开花,沉寂了十几年的网络文学以强硬的姿态闯入人们视野,取得了不俗的成绩。与此同时,随着精品版权运营的快速发展,盗版产业链也逐渐形成,从源头的网文生产,改编制作,甚至衍生作品的呈现,都能看到盗版的身影,侵权技术不断更新,手段也更加隐蔽。

对于网络文学来说,盗版对其的打击是致命性的。不是每一位作家都是辰东、猫腻、唐家三少、天蚕土豆这类大神级的作者,不是每一部作品都像《三生三世十里桃花》《择天记》《琅琊榜》一样受到编辑、导演、受众的青睐,对大部分作家来说,很难达到唐家三少年收入 1.22 亿的水平。更多的作家处于金字塔的底层,靠着写文赚着微薄的稿酬用以生活。即使作家因为作品受到编剧等人的青睐,可以额外获得改编收入,但是影响力大作品反而更容易被盗版盯上。正版版权的蛋糕正在被盗版不断瓜分,严重影响了作家和版权所有者的经济利益。

庆幸的是在大数据时代,随着计算机软件和硬件水平的提升,存储数据成本大大降低、存储内容也大为增加,暂时保存在网络中的盗版侵权数据,通过数据分析、数据挖掘、数据追溯等技术,可以有效呈现侵权行为。盗版网站以什么方式、违规刊登多少部作品、吸引多少受众、获得多少点击量等数据信息一目了然,版权所有者可以通过技术手段固定数字证据,维护自己的合法权益;甚至在某些情况下,通过技术手段完全可以屏蔽盗版链接,将盗版扼杀在萌芽中。利用大数据技术进行网络文学版权保护,避免盗版活动发生,减轻经济损失,势在必行。

### (一) 版权数据优势突出

我们生活在一个被数据包围的时代,网络中的每个人都在制造数据和分享数据。网络文学经过近十五年的快速发展,不仅数量上大大增加,IP 质量也有显著提高。与此同时,针对于网络文学及其衍生作品的侵权行为层出不穷,往往正版作品一上线,我们就可以在盗版网站找到相应的资源,严重损害了版权所有者的利益。数字化基因使其在发展中积累了大量优秀作品的数据库以及侵权案例,为版权问题可视化呈现提供了坚实的

基础。

由腾讯文学与原盛大文学整合而成的阅文集团，可以说是当之无愧引领行业发展的龙头老大，旗下拥有起点中文网、QQ 阅读、红袖添香、潇湘书院等网络原创与阅读品牌，公司文学作品储备达 840 万部，其中包括 800 万部原创作品，覆盖 200 多种题材，平台创作字数达到 414 亿字；<sup>①</sup>仅旗下的晋江文学城就拥有风格迥异的在线作品将近 250 万部。2015 年成立的阿里巴巴文学，旗下拥有书旗小说、淘宝阅读、UC 小说等网络资源，其精心培育的网络小说《华簪录》《全职妈妈向前冲》被《人民日报海外版》点赞……

2017 年引发收视热潮的前三名，无一不是从网络文学改编而来：《三生三世十里桃花》改编自唐七公子《三生三世十里桃花》，《楚乔传》改编自潇湘冬儿《11 处特工皇妃》，《择天记》改编自猫腻《择天记》。其中，电视剧《三生三世十里桃花》官微首播一年之后所发布的微博，4 天之内就有近 3000 个人点赞，近 900 百人评论；动漫《择天记》第 1 集以突破 1500 万点击的成绩创下当时腾讯视频单集动画的最高纪录；手游《花千骨》首月月流水将创新了记录，将近两亿；唐家三少的作品《斗罗大陆》目前已被改编成网页游戏、漫画和动画，漫画改编获得了单本近百万册销量的佳绩……

当你在网络上阅读书籍时，当你在网站上复制转载链接时，所有网络中的动向都留下了痕迹，你看过的小说，看过的章节，以什么样的方式、在哪些地方上传这些作品……通过对用户个人和盗版网站的数据挖掘，深入汇总分析，有利于记录侵权信息，明确侵权的方式，采取相应的措施避免该行为再次发生。

虽然目前硬件条件和技术水平可能达不到对全部存储数据进行分析挖掘，但我们每个人所制造的数据并不是没有用的，我们早已被纳入了网络文学这个平台当中了。网络文学数十年来积累的丰富数据资源及案例信息，为侵权问题的可视化呈现奠定了基础。

## （二）版权威胁超前预判

单纯的数据并没有什么作用，但是当我们用一系列大数据技术分析挖掘之后，汇总的数据就成了有用的信息，我们不仅可以总结出过去的经验，最重要的是可以通过数据预测未来趋势，甚至可以提前预测版权威胁。

Netflix 的《纸牌屋》就是充分利用用户观影数据的典型。在发现英国广播公司经典老片《纸牌屋》很受欢迎，而且这些观众不仅喜欢看著名导演大卫·芬奇拍的电影，

<sup>①</sup>A5 创业网.同为网文大鳄 中文在线和阅文各自讲着怎样的故事？  
[EB/OL].<http://www.admin5.com/article/20170809/773360.shtml>,2017-08-09.



很多还是奥斯卡影帝凯文·史派西的忠实粉丝；与此同时，Netflix 通过数据还发现，每晚守着电视观看节目的人越来越少，传统美剧的周播形式已经不再适合现代人的生活节奏；于是 Netflix 重新拍摄了《纸牌屋》，并选择一次性播放完所有剧集……主观臆断可能会导致我们失误，但是只要我们找到正确的、有用的数据，找到数据之间的关联，一切问题就能迎刃而解。《纸牌屋》不仅接连斩获第 65 届艾美奖单镜头剧集最佳摄影、剧情类剧集最佳导演和剧情类系列剧最佳选角三项大奖，还引发全球观剧热潮，获得口碑和收视率的双丰收。

利用数据进行预测，除了能让我们更好的开发版权之外，还能帮助我们检测未知威胁，将其扼杀在萌芽之中。在把近期美国人最频繁检索的词条和美国疾控中心在 2003 年至 2008 年期间季节性流感传播时期的数据对比之后，谷歌公司的工程师们发现得出的预测与官方数据高度相似，“判断流感从哪里传播”这一结论远远早于官方数据。虽然这只是医疗行业，但是对于网络文学以及衍生作品版权开发和保护同样有着重要意义。

通过数据，我们可以明确感知，哪些点击率较高、易受观众喜欢的作品容易被盗版网站盯上；网络作品主要是通过哪些渠道被人非法复制转载，其主要被哪些网站或者贴吧侵权；在持续开发 IP 的过程中，哪些环节容易产生漏洞，被人利用。我们大胆假设一下，如果能够共享互联网上所有的数据，我们能不能提前发现哪些人浏览或者购买了写文神器，哪些人购买了机器抓取的设备，哪些人在网络上联系手打群体……提前有针对性的进行监控，用邮件或者短信等方式提醒或者警告对方，将盗版扼杀在萌芽里。当然，目前来说，我们的技术还远远达不到预期水平，但是在国外，利用数据技术预测哪些街区、哪些时间段容易发生犯罪行为已经成为现实，在不久的将来，预测版权犯罪将成为可能。

网络作家辰东的《圣墟》、骷髅精灵的《斗战狂潮》、耳根的《一念永恒》等霸占了起点多个榜单，点击量超过千万，本来是应该被重点保护的作品。但可惜的是，当我们在网上搜索其名字时，多家盗版网站堂而皇之地刊登着它们的盗版章节，甚至用浏览器搜索时，首先出现的就是盗版网站的信息，大量正版读者流失到盗版网站，盗版阅读、观影等正成为一种常态。

作为一家监测盗版侵权信息的公司，天擎华美仅在北京就铺设了 100 多台服务器，在海量数据中追踪起盗版电影的路线图，通过算法设置，加上人工识别等辅助手段，寻

找盗版网站的“罪魁祸首”。但是，这也仅仅是针对上映电影的监测平台，如果能建立起一个统一的、专门的针对于网络文学的数据监控系统，那么一旦网站上的文章遭到人工复制转载或者机器抓取，后台服务器就可以直接发出警报，截断非法信号，甚至明确定位出侵权服务器的位置，从源头上打击盗版行为，这样盗版的频率将会大大减少，版权所有者的利益也会得到有效保护。

### （三）全面的一站式维权

虽然目前大数据技术还有待提高，但是我们早已开始利用版权追踪技术来保护作家和原创网站的合法权益。无论是剑网行动，还是盛大的版权追踪系统，它们在对网站上小说特征值进行提取之后，可以与网络上海量的数据进行全天候实时性的对比分析，能够轻易发现并记录网络中出现的盗版信息，开展一站式维权。

不同的数据内容指纹差别其实很大。作为国内网络文学作家的龙头，盛大文学创新院研制出了以文学作品数字指纹技术为核心的版权追踪系统，充分利用了自身的数据资源进行对比，可以瞬间获悉自己作品在整个互联网上的盗版情况：平台作品的哪些章节被人篡改后直接使用，哪些网站刊载有盗版的文章，大概有多少部盗版作品，网站依靠这些作品获得多少点击率并计算出盗版网站相应获得广告收入是多少，甚至可以找出盗版网站服务器的具体分布地点……一旦发现盗版，阅文集团就可以及时通知盗版网站下架抄袭作品，也可以及时通知网站删除盗版链接，避免部分网络服务商以避风港原则逃避责任。

除此之外，阅文集团还可以实时监控盗版作品被删除的情况，对于拒绝删除盗版链接的网站管理者，阅文集团设有专门的律师团队，一旦通过数据检测评估达到立案标准，立即提起诉讼，维护作者和平台的合法权益。而详实的数据侵权信息，为保护自身版权提供了最有力的证据。通过版权追踪系统，盛大文学发现并验证盗版网页高达上千万个，从悠悠书盟、冠华居小说网、小说 520、若雨中文网到阿里旗下的一搜网等等，阅文集团对于盗版网站的打击行为从未停止过。

其实，为了避免侵权而引起相关的法律诉讼，部分网络服务商们也采取了一定措施来阻止盗版的泛滥。全球最大视频共享网站 Youtube 也采用了一种基于数字指纹的识别技术的反盗版过滤系统，它可以在短短几分中内检测出上传的盗版视频片段，发出相应

的警报或者删除盗版视频。<sup>①</sup>在中国国际知识产权维权监测平台上, 已经可以实时记录热门影片的盗版源头。相信在不久的未来, 随着版权追踪技术的进步和硬件设施的提高, 大数据在网络文学版权保护方面将发挥更大的作用。

---

<sup>①</sup> 网易新闻.Youtube 宣布启用一种全新反盗版过滤系统  
[EB/OL].<http://news.163.com/07/1017/09/3R0E1G7I000120GU.html>,2007-10-17.



## 二、大数据时代网络文学版权保护的新问题

从文学产生之初，抄袭就一直存在了，更不用说在网络文学领域了。对于作家来说，作品人物的创作过程，或多或少都存在着自己的影子，他们把自己的人生观或世界观映射在自己所创造的世界中，文学作品像是自己的孩子一样。然而，对于抄袭者来说，作品只是他们赚钱的工具，只要能获得流量，只要能成为“大神”，他们抱着侥幸的心理，打着借鉴的幌子，明目张胆地抄袭他人文章。数据在为他们的“借鉴”提供便利的同时，因为所有的盗版书目、所吸引的受众总浏览量，日访问量甚至有偿广告获取经济利益都可以通过网站的数据进行查询，也为制裁侵权者、保障作者和正版文学网站的利益固定证据。然而在某种程度上，数据虽然帮助我们打击盗版网站，但是也间接加大了我们行政执法的难度，给网络文学版权保护出了一些难题。

### （一）文学生产：内容抄袭与非法改编

#### 1. 内容抄袭范围扩大

在互联网发展初期，抄袭者可以用手打或者复制粘贴的方式拼凑出一篇文章，但他们所抄袭的文章仅仅局限于自己接触到的、阅读过的书籍。在大数据时代，抄袭者本人花费大量人力物力去寻找相关书籍，以复制粘贴进行“写文”的方式早已被时代淘汰。现在只需要一个“写作神器”或者“网络小说写作软件素材”等数据软件，如图 2-1 所示，简单输入几个关键字就可以自动生成一篇文章，将抄袭者从“繁琐”的工作中解脱出来。本来是帮助作者获得写作灵感的数据软件，却成了抄袭神器，抄袭范围扩大了。

据软件业内人士介绍，大部分写作软件由素材库和自动写作系统构成，它们挂着网络小说写作素材软件的旗号，根据作者需要，只需要输入几个关键字就能自动生成一篇几千字的文章，不仅自动生成人名、地名、服饰、人物性格等，乃至故事的情节结构、矛盾冲突、具体的场景都能自动呈现。也就是说只要你有大概的思路，从文章的主体框架到具体的情节描述，软件完全可以代劳，不仅日更字数超额达标，拼凑出来的文章质量也相对较高，自然而然就成为了“写作大神”。笔者在淘宝上搜了一下，“写作神器”价格从 7.99 元到 300 元不等，相对于写文获得的收益而言，价格可谓是十分便宜。

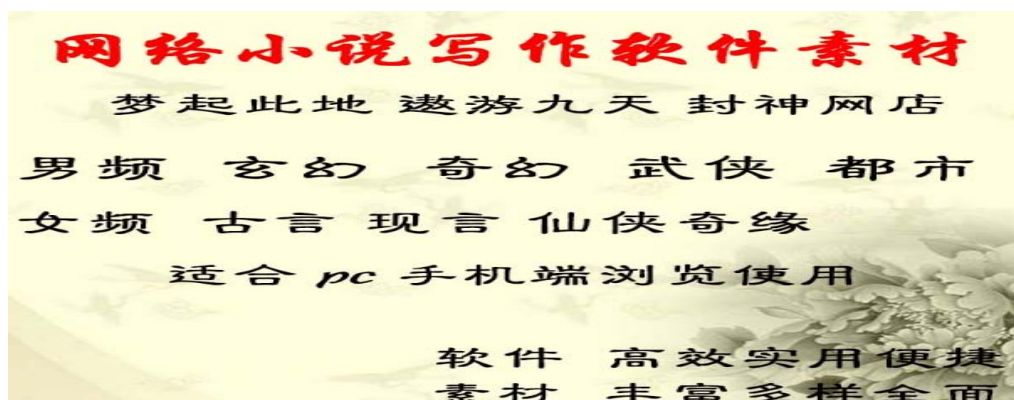


图 2-1

软件库里的素材接近接近千万部，从世界名著到杂谈随笔，应有尽有。输入几个关键词之后，数据软件可以为一个场景提供多种描述方式，供抄袭者选择，因而文章抄袭质量相对较高。从故事主要情节、打斗的场景、人物关系到身穿的服饰、使用的器物……只有你能想到的，没有你抄不到的，由于不是集中抄袭某一部作品，这些隐藏在众多数据中的抄袭章节，很难被人发现其中的端倪。

创下了目前已知的抄袭之最的《庶女有毒》，可以说是利用抄袭软件写作最为知名的作品之一。2013 年，秦简（化名）发表于潇湘书院的《庶女有毒》，一经推出，就成为当时的热门文章，吸引了大批读者为其宣传造势，但也正是因为此书过于热门，被吸引过来的一些常年混迹网文的读者发现了其中的“端倪”。《庶女有毒》正文部分有 294 章，加上番外总共 305 章，题材虽然新颖，但部分章节总给人一些似曾相识之感，在作者驳斥了抄袭的言论之后，网友制作了内容比对的“调色盘”，如图 2-2 所示，发现全文 294 章仅有 9 章未抄袭，整章大范围抄袭可谓常态。从传统文学抄到了网络文学，从武侠奇幻小说抄到了言情小说，从古代小说抄到了近代小说，连天涯、豆瓣等文章也没能逃过其毒手，文中不断浮现《梅花烙》《西厢记》《红楼梦》《九州缥缈录》等文的身影，让人“叹为观止”。

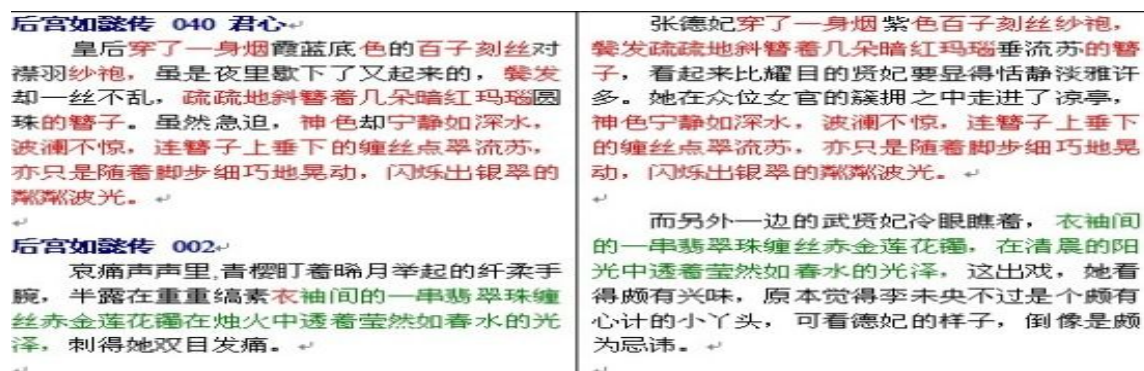


图 2-2

有的网络作家像秦简一样涉嫌抄袭多部作品，也有的实体小说家涉嫌抄袭多部网络作品。可悲的是，作为一名演员，石天琦并不是因为她演出角色让世人铭记，而是因为她的抄袭行为。除了她与某小说网站里应外合偷窃网络作家言晓川 2012 年写的小说《绝色惑妃：太子别嚣张》之外，她的另一本实体小说《水风空落眼前花》，也涉嫌抄袭网络作家冥灵、暗、goodnight 小青、寐语者、秦淮旧梦等 20 多位作家多部作品。

如果不是作品抄袭痕迹过于明显，如果不是作者的“强盗”行为引起众怒，以人工筛查的方式找到网络小说的抄袭部分绝非易事。相比之《水风空落眼前花》，《庶女有毒》并没有原封不动的照搬文章内容，秦简改换字词，调换句式，使拼凑的文章显得更为精妙。借此，她不仅理直气壮的否认自己存在抄袭，其发文平台潇湘书院更是公开表示维护秦简。“你敢保证你的作品里，每一句话都是你自己首创的么？你敢保证在百度绝对找不到和你任何一句话类似的内容么？”潇湘书院这句话将处于灰色地带的抄袭行为以一种血淋漓的现实展现出来，原创作者的利益被本应该保护自己的发文平台网站践踏。

在文章存在涉嫌抄袭可能性时，一句不清楚是否抄袭，华策旗下的克顿影视购买影视改编权并将其改编为《锦绣未央》，其双台收视率破二；在《锦绣未央》作者秦简被 11 位原创作者告上法庭要求其停止侵权、赔礼道歉，并赔偿经济损失的时候，即使北京朝阳区法院已经立案，我们仍然可以在看书堂、33 小说、166 小说网等看到完整版的《庶女有毒》，我们依然可以在腾讯、爱奇艺等平台观看《锦绣未央》，原创作者的利益何在？

现在的网络文学跟 5 年前相比，数量足足多了几倍，越来越多的作家选择在文学网站上发表自己的文章，一些新人作家横空出世，作品不断创下收藏量、浏览量新高，但同时一些热门作品彼此涉嫌抄袭的新闻层出不穷。《庶女有毒》抄袭《后宫如懿传》，《后宫如懿传》又涉嫌抄袭《斛珠夫人》《正说清朝十二后妃》《红楼梦》《宫女谈往录》《枕中记之青城外传》《柔福帝姬》《寂寞宫廷春欲晚》《玉树琼枝作烟萝》等多部作品。笔者真的很好奇，假如跟《庶女有毒》一样，对每个章节进行做相应的“调色盘”，会不会打破它的抄袭记录，创下一个新的抄袭之最！而这仅仅只是我们发现的抄袭文章，互联网上还有多少涉嫌抄袭文章尚未可知。

我们不知道网站上还有多少作者利用过“写作神器”进行写作，还有多少利用大数据进行抄袭。利用抄袭软件获取素材，不仅容易滋生作者的懒汉心理，甚至导致部分作

者离开写作神器就无法下笔。目前我国法律对这种利用软件大范围小规模抄袭行为还没有明确的规定，也无相应的处罚措施，客观上助长了抄袭者的嚣张气焰。在充分利用大数据的同时，加强法律监管，有助于更好地保护网络文学的版权。

## 2. 同人作品侵权突出

相比较于原因，大数据时代更在乎结果，他们用数据说话，直接告诉你社会的热点所在。写文不再仅仅是凭借作者个人的兴趣爱好，而是紧跟社会热点，从影视综艺，到游戏动漫，乃至明星群体，哪些点击率高，哪些受众的关注度高，同人作者就展开具体写作。社会热点早已成为文章写作的主要内容，成为作者吸引读者的主要手段，同人作品也成为大数据时代最容易侵犯版权的一种文体。

与信息时代不同的是，大数据时代的同人文写作早已不再是单单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”的写作方式。根据社会热点或者热门网文作品，选择多个热门题材，由多个故事拼凑成一篇文章。主角无外乎不停地穿越、重生，每一个故事短则几十章，长则上百章，一旦读者流失了，立马换一个读者感兴趣的故事，换一个有话题度的故事，剧情简单、粗暴，以迎合读者为主。

其实，同人这个概念最早来自于日本，首先出现在漫画领域，随着网络小说不断发展，同人小说开始出现。它利用原作品中的人物角色、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的小说。<sup>①</sup>主要可以分为两类，一种是以小说、影视、游戏等作品为基础。或是仅仅借用下人物姓名、性格，描写一个全新的故事；或是主人公以现代的思维，改变了自己应有的命运，故事的典型事件、核心矛盾与原著大量相似。另外一种真人版的同人小说。以现实生活中名人真事，加上自己主观想象，描写一部符合自己幻想小说，相比之下，这一部分更容易造成对他人，特别是对明星名人名誉权的侵犯。因此，我们这里所说的同人小说，重点指第一类。

大部分同人小说家是出于对作品、对于偶像的热爱才进行写文的，他们希望写出另外一种结局，为自己喜欢的主角加油、改变配角或者炮灰的命运，弥补原著的遗憾。像《火影忍者》《海贼王》《名侦探柯南》《美国队长》等世界级的作品，由于配角众多，来自于世界和各地的读者，在网络上发表同人作品，为自己喜欢的人物改写命运。

<sup>①</sup> 马琳.同人小说的著作权问题研究[D].南京:南京师范大学,2017.





图 2-3

阅文集团旗下的晋江网上，专门为衍生/轻小说设立板块，如图 2-3 所示，其中以 25.66 亿的积分高居全站长生殿榜首是林木儿的《敛财人生》。截止 2018 年 2 月初，全文大概写了 11 个故事，530 多万字，光长评约 92 条，足足有 5 页。对于这一类同人小说家来说，由于剧情复杂、故事多样，他们的文章很难被改编成电影、电视剧，如何吸引读者兴趣，写出刺激读者购买欲望的文章，是这些作家首先要考虑的。因此，他们特别注意读者意见，甚至在开一个新的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。这样不仅提高了读者的参与度，而且使自己的故事富有吸引力。一些作者甚至会故意选择一些热门作品进行二次创作，侵犯原创作者版权是他们很难回避的问题。

从钱钟书先生的《围城》到鲁兆明的《围城之后》，从金庸的小说集到江南的《此间的少年》，同人小说对于原创作者的侵权时有发生。笔者截取了 2018 年 2 月 7 日晋江言情长生殿前 20 名的文章，如图 2-4 (a), (b) 所示，发现其中有 12 部网络小说都标有[综]，只有一部《猎人同人一无处不在的龙套生活》是单纯的以日本动漫来写，大范围相似的情节、人物，一旦日本漫话家富坚义博提出抗议，此书必然会因为侵犯版权而下架。此外，其它 11 本网络小说都是一文多事，借鉴多本书的情节框架，有的甚至仅仅借用了书中的人物形象，以现有的中国版权法很难判断其是否侵权。作为中国同人小说侵权第一案，金庸状告江南的《此间的少年》沿用了其一系列武侠作品中的人物姓名，包括大家耳熟能详的郭靖、黄蓉、杨过、段誉等，认为其已经构成著作权侵权及不正当竞争。仅仅是借鉴了武侠作品的名字，就被金庸先生告上法庭，更不用说中国的大多数同人小说，是在大量借鉴剧情的基础上产生的。

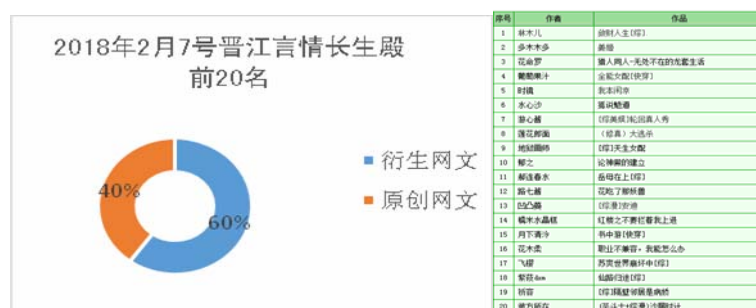


图 2-4 (a), (b)

《哈利波特》的作者罗琳女士在官方网站上虽然明确表示允许读者可以未获授权创作同人作品，但是，她要求同人作品不能用被用于商业用途，也不能有淫移色情的情节；美国迪斯尼为了保护自己的作品，设有专门的律师团队，认为如果没有获得原作者的许可撰写的同人小说，都视为侵犯著作权。晋江网 2 月 7 号上榜的 11 部标有[综]的作品“借鉴”范围从中国古典文学到网络小说，从国内外的热门综艺到日美综漫，更不用说热门的电视剧、游戏了。

只要可以带来人气，带来流量，没有什么他们不可以写的。部分作家从一开始就是凭借快穿起家，例如当天排名第四的作家葡萄果汁，本身没有任何原创作品，无论是正在写的[综]全能女配，还是已经完结的两本作品都是快穿文。这样看来，似乎只要我们掌握了用户数据，就能轻易预测出读者喜欢什么。提前看几篇类似的文章，或者从当前热门的综艺影视剧着手，迎合读者，似乎就很容易成为“写作大神”。一些盗版网站甚至为了增加点击量，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”等各形式的同人作品，大量引用剧情吸引读者。

原创小说网站可以利用大数据指导作者进行写作，帮找到合适的写作方向，促进同人小说商业化，但是这种行为建立在尊重原创作者版权的基础上。我国目前还没有明确的法律规定，出现了管的过严，同人作者不服气；管的松了，原创作家委屈的尴尬局面。

## （二）文学传播：大数据预测作用被盗版网站窃取

### 1. 传播方式：非法复制和盗版链接泛滥

点进去任意一个原创网络小说的网站，哪些是新人榜、哪些是常年榜、哪些是读者呼声最高的、哪些是编辑力推的……只要看到这些榜单，一目了然的数据，在方便读者选书的同时，也使盗文网站明确知道受众的关注点在哪里，知道了机器抓取或者手打的重点在哪里。艾瑞咨询所称的 2016 年全年 PC 端付费阅读收入和移动端付费阅读收入损失，主要是在传播阶段被人侵权所预估的损失，而其主要途径就是盗版链接和非法复制。

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，好事者多名之曰：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	平板电子书
425 第425章 思辨与魔 NEW	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 2-5

非法复制，在网络小说上，即盗版小说网站在未经著作权人的同意的情况下，通过 OCR 文字识别等软件抓取，或者手打等方式将作品或者其中的部分章节复制到自己经营的网站或是贴吧的行为。以晋江文学城的签约作家时镜为例，他的《我 不 成 仙》在修仙类排名靠前且正在连载中。当笔者在 360 搜索上打出“我 不 成 仙”四个字时，首先出现的就是女生文学、晋江文学城、平板电子书三个阅读网站，如图 2-5 所示。在晋江文学城上，作者仍需花钱购买该书的 VIP 章节；然而，当你点开女生文学、平板电子书时，你会发现它们的章节和晋江文学城不仅完全一致而且全部免费，甚至 360 搜索优先提供的是盗版网站的内容。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，大部分读者应该都会选择阅读盗版书籍。

根据艾瑞咨询《2016 年中国网络文学行业研究报告》显示，用户对网络文学版权的认知度偏低，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。<sup>①</sup>当免费盗版章节的视觉效果和正版几乎没有差别的时候，似乎并不难理解看盗版网文受众的心理。除了上文提到的女生文学和平板电子书两个网站之外，针对《我 不 成 仙》小说，笔者粗略统计一下，同一页面的盗版网站还有 2k 小说阅读网、风云小说阅读网、书包网、我 看 书 斋 网 等 4 个 网 站。最为讽刺的是，当笔者点开我 看 书 斋 网 时，在网文章节的最下方出现“如果您发现免费小说我 不 成 仙 最 新 章 节 有 错 误，或发现我 看 书

<sup>①</sup> 艾瑞咨询.2016 年中国网络文学行业研究报告[R/OL].上海:上海艾瑞市场咨询有限公司,2016.03.  
<http://report.iiresearch.cn/report/201603/2540.shtml>.

斋没有更新我不成仙请及时通知我们”的一行小字。一个盗版网站明目张胆的提出希望读者主动提醒他们更新盗文，其猖狂态度可见一斑，而这仅仅是整个网络文学盗版产业的一个缩影。

为了防止盗版，部分作者会根据网站提供的防盗措施采取设置看文防盗比例的方式，比例多设置为 30%或者 50%，订阅低于该比例则在一定时间内不能看文章，如果盗版网站采用机器抓取的方式复制文章，盗版的文章可能出现乱码或者错误。一旦不能看到正版文章，读者就会转战阵地，盗版网站的受众就会开始流失。但是对于不知情的读者来说，看到盗版网站出现相同章节，他们仍然有可能会点进去浏览，即使随后发现文章乱码，退出网页，因此产生的点击量使盗版网站获得了广告收入。贴吧曾经是盗版的重灾区，即使百度曾经下大力气整顿，我们在里面仍然能发现盗文的身影。在《我不成仙》贴吧里，几乎不用花费任何力气，就可以找到其被盗的章节，如图 2-6 所示。部分读者出于分享的态度手打文章，本意并不坏，但是却被一些贴吧利用。为了吸引人气，某些贴吧会组织成员建立专门的“手打团”，边看边打，往往在正版章节出来短短几分钟后，就能在贴吧里找到盗文。为此，有时会出现原创作者对贴吧吧主晚更文章感恩戴德现象。

21	422妖性	mayinjie1...	第422章 妖性 这一刻,见愁是不很说得活出来的。其实若换了一个人在她面前说出这...	儒雅的再...	1-19
2	求第419章节往后的	桃夭1590	求第419章节往后的	你背后有	1-19
9	421	mayinjie1...	第421章 生识不死离 到底是怎么把傅朝生安然从烂柯楼带回来的,见愁已经不想去回...	克拉baby哦	1-19
10	420	mayinjie1...	吓、吓住了。不少人都吓住了,就是崖山这边的自己人都没来由地倒吸了一口凉气。...	师父赐名	1-19
8	419一半	mayinjie1...	第419章 盛怒 真的是说打就打,根本半点犹豫都没有!陆松心里面其实没有多少底。...	师父赐名	1-19

图 2-6

而随着网络技术的发展，网络文学版权开发的不断深化，网络文学已经不在仅仅局限于文本的范畴，涉及到影视、动画、有声读书、游戏等方方面面。以晋江文学城为例，目前已有多位作者授权金达千万级，截止 2017 年 8 月版权成交金额已经破亿元；影视剧有《花千骨》《何以笙箫默》《芈月传》以及即将播出的《知否？知否？应是绿肥红瘦》等等，动漫有《末日曙光》《帝王攻略》《盛世妆娘》《魔道祖师》等等。与此同时，因用户自发上传，网站非法链接转载，第三方运营商审查不严等原因，造成版权所

有者利益受损的情况时有发生。《权利的游戏》在国内播放时，由于网上早已出现带字幕的盗版资源，很多粉丝在国内正版出现之前就看过了，因此压根不会为正版付费；《琅琊榜之风起长林》由爱奇艺独家播放，会员可以提前看六集，然而笔者在某些网站上早已可以找到了和爱奇艺会员一样的集数，也就是说你完全不用购买会员就可以享受会员的待遇；《战狼》《三生三世十里桃花》等电影刚刚上映，网络上就已经出现了多个可以免费下载的盗版链接……

先不说用户用手机浏览盗版文章时，出现先阅读广告再看文章的奇怪景象。部分网站的盗版小说还常常出现章节是对的，内容牛头不对马嘴的场景，上面甚至明目张胆地标注此为防盗章节。这本是作者为了防止文章被盗版，故意发的防盗章节，却被部分网站原封不动地抄了下来，可笑又可恨。对于这些盗版网站而言，只要有人想要免费观看，他们就会提供相应的盗版链接。内容不重要，形式不重要，重要的是利用数据更有针对性地找出较高点击率的作品。

### 2. 传播范围：海外自发翻译侵权

网络小说由于题材多样，人物角色鲜明，故事性较强，说教性弱，更容易被外国读者接受。虽然语言文化不同，但都存在相似之处——真实的人性，再加上跌宕起伏的剧情，风格多样的题材，给读者以身临其境之感，缩小了文化、地域、风俗的差异，得到了很多外国网文读者的认同。网络文学《将夜》《全职高手》等得到了一大批海外读者的喜爱，《三生三世十里桃花》《何以笙箫默》《琅琊榜》等更是获得了一大批死忠粉，从玄幻修仙到言情纯爱，从有地缘优势的东南亚，到美英法等欧美地区，中国网络文学的影响力不断加强。

相比于国内网文的侵权形势而言，中国网络文学译文侵权的问题更加复杂。目前海外读者阅读中国网文的方式主要以海外翻译网站、App、论坛为主，对于网站经营者来说，他们提供热门的网络文学译作并没有收取读者任何费用，主要靠广告、打赏与众筹三种方式提供译文，激励译者更新章节。因而，充分利用现有网站的数据信息，推测出用户喜欢题材文章，十分必要。虽然部分网络小说已经出版海外，但多是以实体书的形式集结出版，读者等待周期较长。为了抢占先机，吸引读者，很多译者事先并没有取得中国作者的授权就已经开始翻译热门文章；与此同时，几个译者同时抢译中国热门网络小说作品的现象时有发生，出现一部作品多个译者的现象。对于那些没有取得授权的网络小说而言，哪个译者先翻译出作品，发表到网上，就能吸引更多的读者进入该网站，

给其带来更多的流量，获得更多的广告收入。为此，部分网站会组织译者对已经翻译的热门作品进行抢译。

其实，早在 2012 年的时候，网络上就已经出现了自发翻译和分享中国网络小说的社区，但当时多数译者处于某种兴趣爱好，而且翻译比较零散，难以长期坚持，直到 Wuxiaworld 等一批海外中国网络小说翻译网站的成立，网络文学翻译才走上了相对正规的道路。美籍华人任我行在看过《星辰变》之后正式进入网文深坑，开始自己创办网站，于是就有了 Wuxiaworld。在 Wuxiaworld 上的作品主要以武侠、玄幻、修仙为主，仅仅两年时间，就拥有大量的人气，日活跃用户数高达 300 万-400 万，其读者更是来自全球 100 多个国家和地区。其它比较知名的国外中国网络文学翻译网站还有 Gravity Tales、起点国际以及 Patreon 等。除此之外，还有以 Novel Updates 为代表汇聚多方内容的流量入口型平台，主要是为了方便读者搜索网络文学小说。

但是在网络文学在进军海外市场取得硕果累累的同时，版权问题也不断凸显，其中最为典型的是 2017 年阅文集团与 Wuxiaworld 的版权之争。其实，Wuxiaworld 自 2014 年创立至今，网站作品主要来源就是起点文学，作为网站创始人任我行翻译第一部作品也是来自起点作家我吃西红柿的《盘龙》。中国网络小说翻译热潮悄然兴起，这虽然在一定程度上扩大了中国网络文学的影响力，但是这完全是在免费的基础上产生的。天下没有免费的午餐，随着起点向国外市场进军，起点国际成立，经济利益的纠纷将原本处于灰色地带的版权问题摆到了明面上，市场上风起云涌，变幻莫测。

起点宣布因为 Wuxiaworld 私自联系 Gravity Tales 的翻译者，并将其未经起点授权的小说移到了自己的网站上，于是正式收回曾经授权给 wuxiaworld 的 20 本小说，并要求其下架 11 本未授权的起点小说。Wuxiaworld 则回应称译者是主动转换阵地申请加入旗下网站开始翻译网络文学，而且曾经与起点达成口头协议（事先在网站上发布新小说的译文，每季度末将再将其纳入协议内），因此才会有新翻译的 11 部起点小说。目前，阅文集团已经与 Gravity Tales（成立于 2015 年 1 月，比 Wuxiaworld 稍晚）确定开展深度合作。

事实的真相我们无法得知，但无论是 Wuxiaworld，还是已经达成合作协议的 Gravity Tales 和起点国际，目前的翻译作品并不收取读者任何阅读费用，几乎都是免费的。Gravity Tales 读者甚至可以和国内读者一样，投票选出自己最喜欢的小说，而得票最高的小说将会获得金钱奖励。起点的目的其实很明显，充分利用自己的版权优势，再将

高质量译者笼络在手中，走付费阅读的道路，进而延伸产业链，将 IP 变现，也许这也是它与提倡免费阅读的 Wuxiaworld 谈不拢的原因之一。

虽然起点作家风凌天下的新作《我是至尊》首次尝试在起点中文网和起点国际同时发布原文和译文，但是这种零时差阅读仅仅是个例，它是建立在风凌天下以往的人气和高质量作品的基础上，并不具有普适性，毕竟原创作家写出 3000 字的文章可能需要花费译者几倍的时间去翻译。如果对于每一本网络作品都实时翻译，先不说将要耗费大量的时间财力，光找到足够的译者就是一件麻烦事。充分利用大数据，找到不同国家和地区人们喜欢的华文网络小说，进而开发相应的衍生作品，也许就可以有重点的对作品进行保护。但是，国内网站网络小说排名的出现使译者对于热门作品了如指掌，尤其在译者待遇普遍底下的网文世界，很多人都是出于分享的目的翻译文章，并没有主动获取作者授权的意识。

实际上，虽然 Wuxiaworld、Gravity Tales 等中国网络文学的英文翻译网站受众看似不少，但是在西方世界还仅仅是小众，很多外国人对此并不了解，他们无法理解中国的网络文学；即使是喜欢看小说的部分外国读者，也曾明确表示，一旦开始收费就不再看文。用户付费意愿整体较低，在这种情况下，如果没有了免费看文吸引读者，那么作品的流传度、影响力便会大打折扣。即便是免费阅读，面对广告收入的诱惑，这些英语网站翻译的小说也开始遭遇盗版链接和非法复制，以达到分流读者的目的。为此，一些译者还专门成立了群组来举报盗版，希望搜索引擎能在搜索结果中屏蔽盗版网站的链接，但是面对大数据的冲击，面对网上海量数据的冲击，单凭几个译者能阻止盗版吗？

### （三）维权面临新困难

#### 1. 侵权作品扩散迅速

当晋江网文学城《我不成仙》还在连载的时候，我们已经可以在十几个网站上找到它目前的所有章节；当《琅琊榜之风起常林》还在爱奇艺会员抢先看时，我们已经可以在好看网上享受爱奇艺 VIP 会员的同等待遇；仅在 2017 年上半年不到三个月时间内，就有上千款涉嫌侵犯盛大游戏著作权的游戏出现……只要正版作品一出现，无论是网络小说还是由此衍生的影视、动画、游戏等，盗版商们会在第一时间采用各种手段“复制粘贴”相应的内容进更新，抢夺受众，甚至出现了各大盗版网站相互“复制粘贴”的乱象。



相对于采用线下出版的文学作品来说，大数据时代，盗版的技术手段不仅更加隐蔽多样，扩散速度更快且影响范围更广，所造成的经济损失更大。对于盗版商而言，只需鼠标轻轻一点，就可以获得自己所需要的所有“资源”，省去了大量印刷、装订、发行等物力和人力费用，即使读者打开网页发现这是错误章节，即使受众瞬间把浮动广告或者弹窗广告关掉，但是这次点击已经完成了，对于盗版网站来说，这是要获得收益的。

对于网站所有者来说，盗版的收益远远大于付出。为此，即使它们明知道自己是违法的，也不惜铤而走险。对于个人而言，只要简单的复制粘贴或者转载，就可以在瞬间向大家分享自己喜欢的文章。在贴吧内只要有“资源”的人下面留言，基本都可以通过邮箱或者百度云盘获得自己想要的，大家“无私”的分享，在客观上助长了盗版的嚣张气焰，作为 IP 开发的源头，网络文学早已不止是文学的范畴，盗版则严重损害了作者的利益和创作热情，最终损害的还是读者的利益。

## 2. 侵权主体认定模糊

数据不一定是信息，但通过对数据的挖掘分析，我们可以得出相应的信息，通过汇总分析挖掘数据，我们可以对侵权的过程进行追溯，找到发生盗版行为的关键所在。但是，当我们被海量的数据包围时，我们的硬件设施和技术手段并没有达到相应的水平，很难去分析所有的数据，这也导致了我们无法充分发挥数据的作用，很难找到所有的侵权主体。

在大数据时代，由于版权生产来源十分多样，对于网络文学而言，原有的作者本位被打破了，读者和作家的界限不再清晰，每个人都可能是网络文学作品的发表者。我们只要伸一伸手指、点击一下电脑，各大网络文学网站的热门小说一目了然，由此引发的大面积抄袭以及同人文的乱象层出不穷。随手分享，便产生了无数的盗版信息源，传输路径像蜘蛛网一样繁杂，凭现有的技术完全无法查出所有的侵权人信息；网站经营者为了获得经济利益，租用国外的服务器存储内容，用技术手段抓取或复制相关章节或者手打更文的现象更是屡禁不止；网络服务商凭借避风港原则纵容受众上传热门网络小说吸引流量，置作者版权于不顾……

无论是起点中文网诉“云霄阁”网站盗版侵权案、还是百度文库侵权案，基本上可以说是集体和集体的碰撞。百度输了，阿里旗下的一搜网输了，可是对于真正主动上传网络文学章节人的惩罚呢？我们可能凭借数据找到侵权的部分主体，保护作者的合法权益，但是，在大数据时代，成千上万条盗版链接加大了我们处理数据的难度，我们很难



或者过于麻烦去找那些直接侵犯作者利益传送盗版文章、视频、音乐的人。至少目前来说，作者自身很难从海量的数据中发现盗文行为，也很难理清众多转载者之间的关系，难以找到侵权者保护自身的合法权益。

### 3. 原创作者维权困难

#### (1) 侵权网文举报难整改

阅文集团旗下的小说网站每个章节都有相应的举报按钮，虽然版面样式可能不同，但每当读者对某一章节有疑问的时候就可以及时点击举报按钮进行举报。但是在大数据的浪潮中，这样的“人工举报”绝对是一件耗费大量时间和精力事情。更不用说阿里旗下、百度旗下的文学网站多是在作品著作权所有者提供相应的侵权证据之后才会去验证是否存在侵权行为，百度文学甚至需要作者提供书面的通知文件，完全忽略了互联网的便利性，使举报程序更为复杂。

对于盗版网站擅自刊登作者拥有著作权的作品，虽然作者一方面可以通过签约网站的盗文投诉等相关功能，向盗文网站发送删文通知，或者在这些非法转载的网站上与管理员联系，请他们删文，但是现实情况不容乐观。盛大旗下部分网站，明确提示作者联系非法转载网站的管理员时，语气要克制客气，是否删文取决于该网站负责人的法律意识等因素，这似乎在告诉作者对于拒绝删文的盗版网站只能自认倒霉。明明应该是自己版权的维护者，却对部分作家网络作品被非法转载事实视而不见，出现了作者自身与盗文网站协商，对部分盗文网站晚更两小时感恩戴德的怪象。

实际上，为了维护自身的版权利益，阅文集团与盗版文学网站的斗争从未停止，仅冠华居小说网一案，追究其涉及侵权网络小说作品高达 3259 部，但是这是在网站负责人的侵权行为以经达到刑事追诉标准时，才开展的诉讼行为。在贴吧中，有作者吐槽，因盗版网站侵权，短短三十分钟，自己一章节的收益少了四分之三，对于部分作家来说，侵权初期就已经造成了自身大量的经济损失，他们只能通过双更、故意断更、设置防盗比例等方式进行保护自己的权利，对于盗版网站和抄袭作家无能为力。

#### (2) 原创作家诉讼困难，效果差

由于抄袭者使用写作神器等数据软件，如何在海量的数据中找到自身被侵权的证据，对于原创作家来说，更是一件难事。《庶女有毒》涉嫌抄袭的作品多达 200 多部，从 2013 年被曝光抄袭，到 2017 年初正式立案，网友陆陆续续花费将近 4 年多的时间，对全文 294 章用“调色盘”一一进行对应，才找到了所有的抄袭章节，联系了大部分被

抄袭的作家，这才有了 2017 年 11 位作家的联合诉讼。但对于今天的一本网络文学动辄一两千章，动辄几百万的字数，仅阅文集团旗下网站晋江文学城就拥有在线作品 240 多万部，仅以人工查重将花费多长时间尚未可知，这些因素无疑为案件诉讼设置了重重障碍。

号称投资 3 亿的电视剧《如懿传》，原著似乎也涉嫌抄袭，在贴吧、知乎也能找到其部分“调色盘”的对比图，但是对于字数 120 多万的文章而言，用人工查重的方式找到其中涉嫌抄袭的证据十分困难，目前并没有形成完整的证据。与此同时，因为大部分网络作家多是自发性写作，自己本身并没有保留证据意识，一旦需要维权，就陷入了自己证明自己是作者的困境。曾经就有文学网站串通女演员更改签约网络作家发表作品时间将其作品据为己有的事情发生，如果不是盗版网站标明发文的时间，恐怕原创作家就要背上抄袭的骂名。即便签约网站配合调查，通过小说网站调取后台信息的流程也颇为复杂，这些都增加了取证的难度。

按照最新的稿酬付费标准，原创作品每千字仅有 80—300 元。<sup>①</sup>但是，一旦提起诉讼，原告不仅面临着花费大量时间寻找证据，还要为律师费、公证费、版权登记费等付账，即使作者胜诉，对于盗版的网站而言，他们获得经济利润远远超出了在法律强制要求下给予作者的补偿。除此之外，部分耽美作家，因为题材不被主流社会接受，他们不愿意提起诉讼，出现在世人眼前，更何况即使提起诉讼，因为法律并没有明确规定，从耽美文变成言情文难以定论是否算作抄袭，原创作者很难打赢官司，获得补偿。《三生三世十里桃花》的作家唐七公子虽然承认抄袭《桃花债》，但是他自始至终拒绝道歉。司法程序周期过长，花费了大量的精力和时间的作家获得的经济补偿过低，有的甚至得不到抄袭者的一句道歉，越来越多的作者保持沉默，选择向盗版妥协。

<sup>①</sup> 中国经济网.《使用文字作品支付报酬办法》发布 稿酬提高.  
[EB/OL]. [http://www.ce.cn/culture/gd/201409/30/t20140930\\_3629620.shtml](http://www.ce.cn/culture/gd/201409/30/t20140930_3629620.shtml),2014-09-30.

### 三、大数据时代网络文学版权保护的建议

2015 年 12 月，移动端阅读覆盖人数近 1.48 亿、PC 端月度覆盖人数近为 1.41 亿；日均覆盖人数，移动端是 PC 端的近 3 倍，达 3297.5 万人。<sup>①</sup>互联网的普及以及智能手机等手持设备的广泛使用，使我们能够 24 小时不间断地产生大数据，有利于寻找潜在市场，开发相应的产品和服务，也利于优化作品、增强精细化运营能力，甚至在一定程度上能预测犯罪，提升政府社会治理的效率。但是，由于现实数据在完整性、精确度和一致性方面存在着差异，数据的质量较差，企业之间数据合作共享的意愿较低，版权追溯技术并不成熟，利用数据加强对网络文学的版权保护还有很长的路要走。

#### （一）搭建大数据监测平台，实现版权信息共享

独木不成林，单独一个数据可能不会告诉我们重要的信息，但是当多个数据汇聚起来，我们所获得的信息量非常庞大。在阅文集团的云中书城上，你基本可以看到其旗下的所有作品，甚至包括一些合作网站的热门作品；帐号共享目前已经实现，读者在其旗下任意一个小说网站注册过之后，凭借相同帐号，就可以登陆其旗下的其它网站，也就是说读者在阅文集团旗下所有网站上的数据都被统一起来。

阅文集团很早就已开始搭建数据共享平台，但是它毕竟是一个企业，以盈利为目的，它不可能把所有的数据无私地分享给阿里或者百度等公司；即使它可以通过数据找到了盗版的源头，提出抗议，由于不是官方执法人员，可能被部分盗版网站无视，甚至还给了部分盗版网站经营者删除盗版链接，盗文的时间，反而使他们逃避了处罚。对于网络文学来言，往往在盗版章节出现的第一时间，读者就已经阅读过了，对于版权所有者造成的损失已经存在了。

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使用，可以快速进行版权认证、预警、监测和维权等。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，那么就可以将各个文学网站的资源结合起来。在这个平台上，所有的网络小说一目了然，既避免了企业担心数据被其它企业利用的现象，又整合了所有的资源，未来可以瞬间发现上传的盗版文章，直接过滤或删除未经授权的网络小说，说

---

<sup>①</sup>艾瑞咨询.2016 年中国网络文学行业研究报告[R/OL].上海:上海艾瑞市场咨询有限公司, 2016.03.  
<http://report.iresearch.cn/report/201603/2540.shtml>.

不定还可以向论文检测一样，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，节约人工成本。

## （二）提升大数据技术水平，系统追踪维权

如果说充分利用的数据堪比黄金，那么沉睡的数据一无是处，只有通过技术手段对其进行分析挖掘，才能得出有用的信息。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，我们要不断提升大数据数据分析、数据挖掘、数据可视化等的技术水平，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事件就是利用技术设施遏制盗版的典型。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，很多作者写文时由于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 改编的时候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。<sup>①</sup>

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”贯穿版权交易的始终，这使版权所有者可以快速获取固定侵权证据，利于自身维权。其实，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 体系中（基于数字版权唯一标识技术），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，有助于数据确权、数据溯源。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，完全可以实现事先预警，事后快速维权。

<sup>①</sup> 纸贵版权.区块链版权能保护原创吗？它的应用性到底如何？  
[EB/OL].<https://www.zhihu.com/question/58651514/answer/158086555>,2017-04-19.

### （三）增大教育投入，培养大数据版权人才

大数据技术人才是复合型的，不仅仅需要具备数据分析的基本能力，还要熟悉相关的业务，只有这样，他才能了解数据的含义，及时通过数据感知业务的变故，发现异常，及时止损。

阅文集团之所以能够起诉多个盗版网站胜诉，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，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。中文在线的移动出版版权保护技术应用系统和网络侵权追踪平台开发等技术，是花费了5年多的时间，投入了大量的技术骨干研发而成的。创造了《纸牌屋》全球观剧热潮的Netflix，两次设下了百万美元的奖金，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。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，数百万篇文学作品，每年光研发投入就超过5000万元。<sup>①</sup>

人才已经成为了企业发展的重中之重。大数据开启了时代的重要转型，改变了人们的思维模式，对医疗、金融、人力资源、天气乃至预测犯罪有着非常重要的作用，数据科学家随之受到热捧。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等主流统计软件分析软件的技能，现有的人才难以满足市场需求。据全球顶尖管理咨询公司麦肯锡(McKinsey)的一份分析报告显示：预计到2018年，大数据或者数据工作者的岗位需求将激增，其中大数据科学家的缺口在14万到19万之间，对于懂得如何利用大数据做决策的分析师和经理的岗位缺口则将达到150万！<sup>②</sup>

增大教育投入，培养大数据人才适应市场发展势在必行，一方面要创新人才培养模式，培养多层次人才，不仅要鼓励学校设立大数据相关专业，还要通过校企合作，培养市场急需的数据分析、挖掘、可视化等方向的人才。另一方面，制定和优化人才引进机制，完善配套的服务设施，设立相关政策服务，吸引海内外高精尖大数据人才就业创业。

<sup>①</sup> 中国新闻出版广电网.冠勇科技:面向全球提供版权大数据服务  
[EB/OL].[http://www.chinaxwcb.com/2017-08/10/content\\_359452.htm](http://www.chinaxwcb.com/2017-08/10/content_359452.htm),2017-08-10.

<sup>②</sup> 阿里云大学.R语言入门与实战:大数据人才缺口巨大,现在学习还不晚  
[EB/OL].<https://zhuanlan.zhihu.com/p/27460707>,2017-06-19.

#### （四）加强顶层设计，完善相应的法律法规

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。<sup>①</sup>

其实，从《信息网络传播权保护条例》到《关于加强网络文学作品版权管理的通知》，我国制订了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，优化版权环境，维护网络著作权人的合法权益。但是由于法律的滞后性，对利用数据软件的写作行为，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，甚至可能连是否侵权都尚未可知，客观上助长了部分侵权人的嚣张气焰。《庶女有毒》涉嫌抄袭一案，可以说是对于大数据时代下滥用写作软件的一种反抗。政府亟需出台新的法律法规，保护版权所有者的利益。

同时，国家也要从经济和政策上大力支持大数据建设和发展。根据艾瑞咨询的数据，2016年已上市的互联网公司Top3均为数据驱动型企业，分别为阿里巴巴、腾讯和百度，而这正与阿里文学、阅文集团和百度文学一一对应，很难说文学三巨头的发展与数据没有任何关系。其实，无论是从2005年开始已经展开13年的剑网行动，还是2016年建立的建立“黑白名单”制度，都离不开数据的支持。制定大数据发展的相关规划、加强对数据管理，对推动数据共享、形成统一的数据标准对于网络文学的版权保护至关重要。

---

<sup>①</sup> 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].[http://www.gov.cn/xinwen/2017-02/16/content\\_5168508.htm](http://www.gov.cn/xinwen/2017-02/16/content_5168508.htm),2017-02-16.

## 结 语

目前,网络著作权技术保护措施有很多,例如反复制技术、电子水印技术、数字指纹技术、数字作品使用追踪技术等等。但是,由于种种原因,尤其是高昂的成本,大量的资金技术人才投入,让小型的文学网站不愿意采取相关的技术措施去维护自己的版权。在某种程度上,在读者点击率的争夺战上,它们自己不仅是侵权的受害者,同时也是侵权行为的实施者,这加大了利用数据技术进行网络文学版权保护的难度。

当然,大数据也并不是万能的、毫无缺陷的,我们不仅要看到大数据在版权方面的优势,还要警惕它的不足。它虽然可以帮助我们预测信息,但是这只是参考答案,为我们做出决定提供更为丰富的资料信息,并不等同于现实。而且随着不同形式的数据汇总,数据质量的问题越发突出,提升数据的质量,制定相应的数据标准,对于加强网络文学版权保护同样有着重要的作用。

在论文写作过程中,由于涉及大数据、数据挖掘、数据质量、数据分析等很多专业方面的知识,笔者对如何利用大数据进行网络保护的分析还处于一个较浅的层面上,提出的对策可能还有所欠缺,在今后的学习和工作中,笔者会继续努力,丰富自己相关的知识,以期待能提出更好的建议,进一步加强对网络文学的版权保护。





## 参考文献

### 一、专著类

- [1] 杨小兰.网络著作权研究[M].北京:知识产权出版,2012.
- [2] 史辉.网络著作权制度研究:权利保护、限制与交易[M].北京:金城出版社,2016.
- [3] 著作权侵权专题判解与学理研究(第二分册:网络空间著作权)[M].北京:中国大百科全书出版社,2010.
- [4] 张玉宏.品味大数据[M].北京:北京大学出版社,2016.
- [5] 段云峰.大数据和大分析[M].北京:人民邮电出版社出版发行,2015.
- [6] 蔡莉,朱杨勇.大数据质量[M].上海:上海科学技术出版社,2017.
- [7] 张平.知识产权法[M].北京:北京大学出版社,2015.
- [8] 杜启杰.大数据大战略[M].北京:中国言实出版社,2017.
- [9] 汪勇,史学清.网络侵权案例研究[M].北京:中国民主法制出版社,2009.
- [10] 李晶晶.数字环境下中美版权法律制度比较研究[M].北京:人民日报出版社,2016.
- [11] 谭磊.大数据挖掘[M].北京:电子工业出版社,2013.
- [12] 郎为民.漫话大数据[M].北京:人民邮电出版社,2014.
- [13] 维克托·迈尔-舍恩伯格,肯尼思·库克耶.大数据时代[M].浙江:浙江人民出版社,2013.
- [14] 周苏,王文.大数据导论[M].北京:清华大学出版社,2016.
- [15] 欧阳友权.网络文学本体论[M].北京:中国文联出版社,2004.

### 二、期刊类

- [1] 鲍嫻.IP 热潮下的网络文学版权困境探析——以 2016 年三大网络文学版权事件为例[J].科技与出版,2017(04).
- [2] 唐思慧.大数据环境下文本和数据挖掘的版权例外研究——以欧盟《DSM 版权指令》提案为视角[J].知识产权,2017(10).
- [3] 李婷,刘广登.大数据时代数字版权保护体系的构建与完善[J].出版广角,2017(10).
- [4] 李钢.大数据时代文本挖掘的版权例外[J].图书馆工作与研究,2016(03).
- [5] 向鹏,张婷.论大数据时代网络版权保护的刑法治理[J].出版广角,2017(18).
- [6] 王兵,李邈.数字出版领域中大数据技术的应用及前景思考[J].数字技术与应用,2016(02).
- [7] 朱磊,荆磊.“互联网+”环境下数字版权保护模式[J].电子技术与软件工程,2017(20).
- [8] 陈晓林.大数据云监测平台的构建与版权保护[J].内蒙古科技与经济,2016(21).
- [9] 许波.大数据时代数字出版的困境与对策研究[J].出版广角,2017(05).

- [10] 赵丰,周围.基于区块链技术保护数字版权问题探析[J].科技与法律,2017(01).
- [11] 张从龙.坚持自主创新构建国家数字版权保护技术支撑体系[J].出版发行研究,2017(02).
- [12] 赵兵,郭才正,祁汝峰,马云辉,王飞.论“互联网+”时代的网络文学版权保护[J].新闻研究导刊,2017,8(19).
- [13] 李东.数字版权保护存在的问题及解决措施[J].中国报业,2018(03).
- [14] 王喜军,侯铭峰.网络文学版权问题分析[J].西安石油大学学报(社会科学版),2017,26(05).
- [15] 付智慧.我国数字版权应用现状及发展趋势分析[J].产业与科技论坛,2018,17(01).
- [16] 高纯娟.我国网络文学海外译介与输出研究[J].出版广角,2017(18).
- [17] 张明,顾亚男.一种基于数字版权保护的数字盲水印算法研究[J].无线互联科技,2017(23).
- [18] 张振宇,周莉.“大数据出版”的理念、方法及发展路径[J].出版发行研究,2015(01).
- [19] 宋学清,乔焕江.大数据背景下网络文学的新生产机制与新景观[J].文艺评论,2017(03).
- [20] 周平.大数据出版与版权保护初探[J].中国经贸导刊,2016(35).
- [21] 丁莉,张文斐.大数据背景下网络文学生产策略研究[J].长春中医药大学学报,2016,32(03).
- [22] 张金婵,孙丽.大数据如何助推网络文学 IP 发展[J].科技传播,2017,9(13).
- [23] 刘丹柏.大数据互联网技术在网络文学中的应用研究[J].自动化与仪器仪表,2017(08).
- [24] 齐元军.大数据时代数字出版版权保护的难点与策略研究[J].科技与出版,2014(11).
- [25] 陈晓林.大数据云监测平台的构建与版权保护[J].内蒙古科技与经济,2016(21).
- [26] 陈志文.数字出版资源版权保护探析[J].出版与印刷,2017(01).
- [27] 邵琰.数字出版著作权保护策略研究[J].出版广角,2015(14).
- [28] 姚万勤.刑法应如何应对大数据时代数字版权保护的“焦虑”[J].重庆邮电大学学报(社会科学版),2016,28(05).

### 三、报纸类

- [1] 苏墨,兰德华.大数据+网络文学的 N 种可能性[N].工人日报,2017-01-23(007).
- [2] 袁跃兴.大数据背后的网络文学[N].深圳商报,2015-10-8(A10).
- [3] 金晶.版权:站到大数据肩上[N].经济日报,2014-12-2(016).
- [4] 胡建辉.大数据助力互联网版权立体维权[N].法制日报,2014-09-18(006).
- [5] 李明远.冠勇科技:面向全球提供版权大数据服务[N].中文新闻出版广电报,2017-8-10(006).
- [6] 唐婷.数字版权保护借力区块链技术[N].科技日报,2017-03-29(003).

### 四、学位论文类

- [1] 姚婷婷.阅文集团 IP 运营研究[D].南京:南京大学,2016.
- [2] 奥尼巴骆尧.盛大文学诉百度侵权案法律分析[D].兰州:兰州大学,2014.
- [3] 赵怡君.大数据时代媒介及媒介产品的交互影响[D].武汉:武汉体育学院,2015.

- [4] 张新雯.互联网时代的微版权运营模式探析[D].北京:北京印刷学院,2016.
- [5] 张鑫.腾讯文学原创频道研究[D].保定:河北大学,2015.
- [6] 廖宏斌.起点中文网的发展探析[D].成都:西南财经大学,2009.
- [7] 何慧龙.我国网络文学网站产业链发展研究[D].开封:河南大学,2015.
- [8] 余场.媒融合背景下网络小说影视改编剧传播策略研究[D].武汉:湖北大学,2016.
- [9] 郝珊珊.大陆网络文学的十年发展和现实反思[D].福州:福建师范大学,2008.
- [10] 李亮.我国网络文学实体出版中的问题研究[D].保定:河北大学,2015.
- [11] 尤杰.数字传播时代的版权与言论自由权之争:对转换性使用的哲学思考[D].上海:上海交通大学,2011.
- [12] 马琳.同人小说的著作权问题研究[D].南京:南京师范大学,2017.
- [13] 牛敏.基于区块链技术的数字版权管理模式研究[D].北京:北京印刷学院,2017.

## 五、报告类

- [1] 中国互联网络信息中心.第 41 次中国互联网络发展状况统计报告[R/OL].北京:中国互联网络信息中心,2018.01.
- [2] 艾瑞咨询.2016 年中国网络文学版权保护白皮书[R/OL].上海:上海艾瑞市场咨询有限公司,2017.04.
- [3] 艾瑞咨询.2017 年中国网络文学出海白皮书[R/OL].上海:上海艾瑞市场咨询有限公司,2017.09.
- [4] 艾瑞咨询.2016 年中国数据驱动型互联网企业大数据产品研究报告[R/OL].上海:上海艾瑞市场咨询有限公司,2016.12.
- [5] 艾瑞咨询.2016 年网文江湖群英谱——中国网络文学作者洞察报告[R/OL].上海:上海艾瑞市场咨询有限公司,2016.12.
- [6] 艾瑞咨询.2015 年中国网络文学版权保护白皮书简版[R/OL].上海:艾瑞咨询集团, 2016.01.
- [7] 艾瑞咨询.2016 年中国网络文学行业研究报告[R/OL].上海:艾瑞咨询集团, 2016.03.

## 六、电子文献

- [1] 金融界.“剑网 2017”专项行动关闭侵权盗版网站 2554 个  
[EB/OL].<http://finance.jrj.com.cn/2018/01/16155323954837.shtml>,2018-01-16.
- [2] 社微世 Service.我眼中的 IP 发展史:网文作品的兴起史  
[EB/OL].[http://www.sohu.com/a/204094027\\_100049702](http://www.sohu.com/a/204094027_100049702),2017-11-13.
- [3] FineBI.再谈大数据的特征,感受数据之大  
[EB/OL].[http://blog.sina.com.cn/s/blog\\_ecfc26de0102vh3l.html](http://blog.sina.com.cn/s/blog_ecfc26de0102vh3l.html),2015-03-14.
- [4] 腾讯智慧.慕思走近腾讯智慧,碰撞家居品牌营销新火花  
[EB/OL].[https://mp.weixin.qq.com/s?\\_\\_biz=MjM5ODY2ODMyMg%3D%3D&idx=1&mid=2650366124&sn=31e61b05695f81fc5006be6ee7d9010b](https://mp.weixin.qq.com/s?__biz=MjM5ODY2ODMyMg%3D%3D&idx=1&mid=2650366124&sn=31e61b05695f81fc5006be6ee7d9010b),2016-10-31.

- [5] A5 创业网.同为网文大鳄 中文在线和阅文各自讲着怎样的故事?  
[EB/OL].<http://www.admin5.com/article/20170809/773360.shtml>,2017-08-09.
- [6] 网易新闻.Youtube 宣布启用一种全新反盗版过滤系统  
[EB/OL].<http://news.163.com/07/1017/09/3R0E1G7I000120GU.html>,2007-10-17.
- [7] 中国经济网.《使用文字作品支付报酬办法》发布 稿酬提高.  
[EB/OL].[http://www.ce.cn/culture/0gd/201409/30/t20140930\\_3629620.shtml](http://www.ce.cn/culture/0gd/201409/30/t20140930_3629620.shtml),2014-09-30.
- [8] 纸贵版权.区块链版权能保护原创吗? 它的应用性到底如何?  
[EB/OL].<https://www.zhihu.com/question/58651514/answer/158086555>,2017-04-19.
- [9] 中国新闻出版广电网.冠勇科技:面向全球提供版权大数据服务  
[EB/OL].[http://www.chinaxweb.com/2017-08/10/content\\_359452.htm](http://www.chinaxweb.com/2017-08/10/content_359452.htm),2017-08-10.
- [10] 阿里云大学.R 语言入门与实战:大数据人才缺口巨大,现在学习还不晚  
[EB/OL].<https://zhuanlan.zhihu.com/p/27460707>,2017-06-19.
- [11] 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].[http://www.gov.cn/xinwen/2017-02/16/content\\_5168508.htm](http://www.gov.cn/xinwen/2017-02/16/content_5168508.htm),2017-02-16.

## 致 谢

明德新民，止于至善，2011年9月1号，我第一次迈进了河大的校门，那时的我还不懂这句话的意思，转眼间七年过去了，我即将离开河大，这时的我才稍微懂得其中的内涵，以教育来明德，自我革新求进，河大的校训如酒般醇厚，历久弥香。

与本科毕业时候的离别不同，因为当时已经确定在母校读研，处于毕业季的我似乎并没有那么多的离愁别绪，唯一伤感的是与好友的分别。三年后，当自己真的要离开河大，离开开封的时候，却发现自己有那么多的不舍。忘不了第一次见到清明上河园的喜悦，也忘不了铁塔湖粼粼的波光；忘不了方正威仪的城墙，也忘不了气势恢宏的大礼堂；忘不了河大和蔼可亲的老师，也忘不了河大西门可口的小吃……

不知不觉我在河大已经呆了七年，从仁和公寓转到校内，从编辑出版换到了新闻学。这里曾是我梦想开始的地方，也是我梦想即将启航的地方。开封，这座城市少了些大城市的繁华与浮躁，却多了些悠闲与逍遥，校园里学生的书生气也似乎更为浓厚，也许这也就是老师曾经说的“铁塔牌的孩子们比较内秀”的原因吧。

很庆幸，在我本科期间，作为一名志愿者为河大的百年校庆服务；很珍惜，我参加了新闻与传播学院的十年院庆；很开心，编辑出版专业的“书是生活”是从我们这一届开始的……在这里，我有太多太多的感动，有太多太多的回忆。在这里，我收获了太多的良师益友，每当困顿难行之时，总会有他们相伴左右为我指点迷津、排忧解难。

我特别感谢我的导师，王志刚老师。研究生三年期间，王老师不仅在学术上常常指点我，也教会我很多做人的道理，对我的帮助很大。在整个论文的写作过程中，王老师也提了很多宝贵的意见，从论文开题到框架的完善，从文章大体方向思路到具体的细节，始终耐心地纠正我的错误。我也特别给感谢李建伟老师、杨萌芽老师、郭奇老师、王鹏飞老师、段乐川老师、许莹老师、于春生老师、祁涛老师、刘杨老师在论文写作过程中给我提出的意见建议，正是在他们的指导帮助下，我的论文才能进一步完善。

我还要感谢我的朋友们，正是因为你们的鼓励帮助，正是因为你们的陪伴，我才坚持走下去。谢谢你们！

度冉

2018年4月18日



## 攻读学位期间发表的学术论文目录

- [1] 王志刚,度冉.华文出版“走出去”战略及实施策略探析[J].出版广角,2016(277).
- [2] 度冉.浅谈媒体在医患纠纷中扮演的角色[J].传播与版权,2016(40).



民善  
新至  
德於  
明止