

ISSN 1009-3125 CN11-4099/G2

全国中文核心期刊

CSSCI 来源期刊

中国优秀图书馆学期刊

中国人文社会科学核心期刊



Journal of
the National Library of China

國家圖書館學刊

任繼愈題

主管单位：中华人民共和国文化和旅游部

主办单位：中国国家图书馆

2022年第1期

双月刊

主管单位：中华人民共和国文化和旅游部

主办单位：中国国家图书馆

主 编：熊远明

常务副主编：陈清慧

编委（以姓氏笔画为序）：

卢海燕 杨沛超 吴建中 汪东波 沈国朝

张晓林 陈 力 陈传夫 陈清慧 柯 平

顾 犇 蒋永福 熊远明

编辑部成员：陈清慧 张收棉 刘冰雪

本期责编：刘冰雪

编辑出版：《国家图书馆学刊》编辑部

地 址：北京市海淀区中关村南大街33号

邮 编：100081

电 话：010-88545737

网 址：<http://gtxk.nlc.cn>

承印单位：北京科信印刷有限公司

国内发行：北京市报刊发行局

国外发行：中国国际图书贸易总公司

(中国国际书店北京399信箱)

JOURNAL OF THE NATIONAL LIBRARY OF CHINA

Vol. 31, No. 1, Feb. 2022

Supervised by: Ministry of Culture and Tourism of
the People's Republic of China

Sponsored by: National Library of China

Edited by: Editorial Office of Journal of the National
Library of China

No. 33 Zhongguancun Nandajie

Haidian District, Beijing 100081, China

E-mail: gtxk@nlc.cn

Website: <http://gtxk.nlc.cn>

Distributed by: China International Book Trading
Corporation

目 录

专家视点

数字公平实现进程中的外部效应研究 闫 慧 003

政策与法规研究

《公共图书馆法》表达与实践的实证研究 邓杰明 013

中国图书馆学学科建设

图书馆学专业教育发展的成就、隐忧与出路 龚蛟腾等 022

以基础知识领域推动图书馆学课程改革

——国际图联《图书馆与信息学专业教育方案指南（2021年草案）》

的译读与启示 张晓芳 王凤蛟 033

管理与服务

公共图书馆志愿服务研究：起源、类型与创新进展 金武刚等 042

国外公共图书馆重大突发事件应急管理机制案例研究 刘 冰等 052

困境与破局：文旅融合背景下的公共图书馆旅游创新研究

李泽华 065

生命周期视角下“一带一路”专题数据库利益相关者分析

何 依等 074

公共图书馆老年人健康信息服务策略研究 刘一鸣 朱萍萍 084

英美大学图书馆学生健康服务研究 黄红华 杨晓平 095

智慧图书馆

我国智慧图书馆文本数据挖掘侵权风险与治理对策研究

闫宇晨 106

业界动态

2020年度高校图书馆基本数据排行榜发布(021) 美国国会图书馆发布
2022—2026年数字馆藏发展战略(064)

JOURNAL OF THE NATIONAL LIBRARY OF CHINA

Volume 31 Number 1

February 2022 Serial No. 139

Contents

Expert View	
Study on Externality in Process of Digital Equity	Yan Hui 003
Research on Policy and Regulation	
Empirical Research on the Expression and Practice of <i>Public Library Law</i>	Deng Jieming 013
Discipline Construction of Library Science in China	
Achievements, Plights and Outlet of the Development of Library Science Education	Gong Jiaoteng, et al 022
Curriculum Reform of Library Science Driven by the Basic Knowledge Areas ——Interpretation and Enlightenment of IFLA <i>Guidelines for Professional Library and Information Science Education Programmes (2021 Draft)</i>	Zhang Xiaofang, Wang Fengjiao 033
Management and Service	
Study on Volunteer Service in Public Libraries: Origin, Type and Innovative Progress	Jin Wugang, et al 042
Case Study on Emergency Management Mechanism of Major Emergencies in Foreign Public Libraries	Liu Bing, et al 052
Dilemma and Breaking: Research on Tourism Innovation of Public Library under the Background of Culture-Tourism Integration	Li Zehua 065
Stakeholder Analysis of the “Belt and Road Initiative” Thematic Database from the Perspective of Life Cycle	He Yi, et al 074
Research on Health Information Service Strategy for the Elderly in Public Library	Liu Yiming, Zhu Pingping 084
Study on University Library Students’ Health Services in the United States of America and United Kingdom	Huang Honghua, Yang Xiaoping 095
Smart Library	
Countermeasures for Infringement Risk Management in the Mine of Text Data in Chinese Smart Libraries	Yan Yuchen 106

我国智慧图书馆文本数据挖掘侵权风险与对策研究

闫宇晨

摘要 数字时代下,文本数据挖掘技术的广泛应用对我国智慧图书馆的建设意义重大。但文本数据挖掘是否属于著作权例外这一问题引起国内外学界、实务界广泛争议,也给图书馆新技术的应用带来了一定的侵权风险。本文对近年域外相关制度改革进行梳理与反思,发现相关做法不具备法律移植的必要性与可行性;对文本数据挖掘的行为本质进一步分析发现,可以在著作权法框架下寻求一种“二分式”的侵权判断法,正确认定我国图书馆文本数据挖掘行为合法性问题,从而推动数字经济发展。表1。图1。参考文献24。

关键词 智慧图书馆 文本数据挖掘 著作权例外 合理使用

Research on the Infringement Risk and Countermeasures in the Text Data Mining of Smart Libraries in China

Yan Yuchen

Abstract: In the digital age, the wide application of text data mining technology is of great significance to the construction of smart library in China. However, whether text data mining belongs to copyright exception has aroused widespread controversy in academic and practical circles at home and abroad, and also brought infringement risks to the application of new library technologies. This article reviews and reflects the relevant extraterritorial system reforms in recent years, and finds out that the relevant practices do not yet have the necessity and feasibility of legal transplantation. A further analysis of the essence of text data mining found a “dichotomous” infringement judgment method within the framework of the copyright law, which can correctly identify the legitimacy of the text data mining of libraries in China, thereby promoting the development of digital economy. 1 tab. 1 fig. 24 refs.

Keywords: Smart Library; Text Data Mining; Copyright Exception; Fair Use

1 问题的提出

随着计算机技术的快速发展,大数据分析 with 深度学习等新技术将会在智慧图书馆建设中得以广泛应用,人工智能借助机器分析可以从庞大的数据中挖掘出各类有价值的信息,其中,从自然语言文本中挖掘用户所感兴趣的模式与知识的技术,一般称为文本数据挖掘(Text Data Mining, TDM)^{[1](1-2)}。TDM在国民经济各个领域都具有广泛的应用前景,尤其在公共文化服务领域,传统图书馆在新技术的推动下朝着智慧图书馆发展,TDM所带来的精准服务与智能化服务将进一步促进公共文化服务领域的繁荣。但随着我国知识产权保护力度的不

断加大,TDM著作权侵权风险给智慧图书馆建设与发展带来了不小的挑战^[2]。一方面,以公共图书馆为代表的公共文化服务行业主张应当将TDM纳入著作权例外的范畴,力图借助自身行业特点及优势提高信息资源利用率,实现公共利益最大化;另一方面,传统著作权产业则坚持任何传播技术的应用,必须给著作权以充分的尊重,作者通过授权许可的方式实现创作的对价是使用作品的必备前提^[3]。

近年来伴随着英美、日本、欧盟等国家和地区各自TDM著作权例外制度的构建,我国学界就是否应当对相关制度进行法律移植,以及如何改造本土合理使用制度以适应数字经济的发展展开了

激烈的争论。归纳看来,主要包括以下两种意见:一种意见认为,应借我国著作权法第三次修订之机在合理使用制度中增设 TDM 著作权例外条款,即“出于科学研究或其他合理目的,可以在必要限度内使用已经合法接触的作品开展信息分析”^[4]。如此,以增进社会福祉为主要目的的公共图书馆,其 TDM 行为可以在此条款支撑下合法开展“挖掘”活动。另一种意见认为,我国合理使用制度封闭式列举的立法模式过于僵化,弊端凸显,难以满足数字技术快速发展的需求。应在合理使用制度中引入“一般条款”,参考借鉴美国“四要素”判断法,为转换性使用的运用提供充足的空间。实践中,可将能够产生独立数据价值的机器分析行为认定为转换性使用,以此作为合法性基础激励创新发展^[5]。但就法律修订的实际情况来看,2020 年 11 月修订通过的《著作权法》中与图书馆合理使用密切相关的第二十四条未作大幅修改,并未就此问题作出学界期待的回应。那么,如何在现有著作权法框架下有效应对图书馆 TDM 侵权风险,维护著作权人利益的同时推动智慧图书馆建设是我们当下无法回避的问题。

2 对近年国外 TDM 著作权例外制度改革的梳理与反思

2.1 近年国外 TDM 著作权例外的制度变革

面对大数据时代 TDM 的著作权困境,各主要发达国家都做出了一定的制度性回应,大致可分为两种做法。第一种以美国为典型代表,其依托本国深厚的判例法基础,借助合理使用的“一般条款”来判断 TDM 是否属于著作权例外,也即合理使用的“四要素”判断法:第一项要素,使用行为的目的与特性;第二项要素,作品的性质;第三项要素,使用作品的比重;第四项要素,对作品潜在市场所造成的影响^①。实践中,法官通过对

“四要素”进行判断,决定 TDM 行为是否增加了原作品新表达、新意义或新功能,进而认定是否构成以转换性使用(Transformative use)为目的的合理使用^②。总的来看,美国版权法是以一种开放的态度对待新技术背景下作品使用方式的变化,依靠法官逐案判断的方式认定某种 TDM 行为是否符合著作权法激励创新的立法宗旨,转换性使用的判断在实践中也充分展现了其灵活性。

第二种以日欧等发达国家、地区为代表,他们通过专门立法的方式规定 TDM 著作权例外。2009 年日本著作权法进行修订,首次将计算机信息分析纳入 TDM 著作权例外,该法第 47 条规定合法获取作品的行为主体可以信息分析为目的实施复制、改编行为,并强调 TDM 著作权例外控制在仅用于计算机信息分析的必要限度内,但并未排除商业目的的 TDM^[6]。2014 年英国对《版权、设计与专利法 1988》进行修订以回应数字时代新形势下信息技术发展的需要,新增 TDM 著作权例外规定,任何人都可以在无须征得版权人同意的情况下出于非商业目的开展 TDM 活动。该法案第 29 条 a 款明确了构成 TDM 著作权例外的要件:第一,行为人必须出于非商业目的;第二, TDM 涉及信息不能转让他人或用作信息分析以外的目的;第三,任何意图阻止或限制 TDM 的合同条款都不具有可执行力^[7]。2018 年颁布生效的德国《著作权与邻接权法》第 44 条 b 款和第 60 条 d 款规定,在权利人未声明保留前提下,科研机构出于非商业性研究目的复制作品属于 TDM 著作权例外,此外,为合作进行科学研究或为审查科学研究的质量而将 TDM 成果传播至特定第三方也可纳入例外范畴^[8]。2019 年 4 月欧盟正式批准实施《数字化单一市场版权指令》(以下称《指令》),《指令》第 3 条规定了“科研目的下的 TDM 例外”,即科研机构在科学研究目的下开展 TDM 而实施的复制与撷取行为不侵犯著作权或数据库权,原则上权利人不得通过合同条款规避

① 17 U. S. C. § 107。

② Campbell v. Acuff-Rose Music, Inc., 510 U. S. 569(1994)。

TDM,需指出的是,科研机构可以与私人主体合作进行 TDM 并适当兼顾商业目的。《指令》第 4 条将进一步将非科研目的主体也纳入 TDM 著作权例外

外行为主体范围,但该主体只能进行临时复制行为,且允许权利人通过合适的方式排除 TDM^[9]。

表 1 国外主要国家 TDM 著作权例外制度梳理

国家或组织	改革法案	行为主体	行为内容	行为目的	特殊要求
美国	《版权法》第 107 条	合理使用“四要素”判断法			无
日本	《著作权法》第 47 条	未作限定	复制、改编行为	信息分析(未排除商业目的)	必要限度内的信息分析
英国	《版权、设计与专利法 1988》第 29 条 a 款	未作限定	信息分析	仅为信息分析且非商业目的	通过合同非法排除无效
德国	《著作权与邻接权法》第 44 条 b 款、第 60 条 d 款	科研机构	复制	科研目的	权利人未声明保留
欧盟	《数字化单一市场版权指令》第 3、4 条	科研机构	复制和撷取	科学研究(可兼具商业目的)	通过合同非法排除无效
		非科研机构	临时复制	商业目的	可通过合适方式排除

总的来说,世界各主要发达国家已充分认识到 TDM 所带来的经济机遇,并积极探索以立法的方式降低数据挖掘成本、提高数据挖掘效率。但可看出,国外 TDM 著作权例外改革展现出多元化的特征,行为主体、行为目的、行为内容等关键要素各国规定差异化明显,并未形成统一做法(表 1)。归纳看来,分为两种立法模式:一是增设 TDM 著作权例外条款;二是将 TDM 纳入合理使用判断,以“一般条款”认定其是否构成转换性使用。我们需要进一步思考的是,这两种立法模式是否具备法律移植的必要性和可能性,能否用于解决我国 TDM 著作权侵权问题。

2.2 对国外 TDM 著作权例外制度改革的思考

2.2.1 贸然增设 TDM 著作权例外易引发实践混乱冲突

各国 TDM 著作权例外制度仍处于探索建立过程中,不同国家展现出较大的立法差异,现有立法成果存在概念范畴不清、适用规则不明的问题,

尚缺乏可借鉴的成熟规则体系。其一,概念范畴不清。以欧盟《指令》中“科研机构”这一主体概念为例,《指令》第 2 条将“科研机构”解释为具有公共主体地位的科研单位和文化遗产单位,前者主要指大学(及其图书馆)与科研院所,后者是指图书馆、档案馆、博物馆、影音文化遗产机构。然而,考虑到实践中 TDM 较多采用公私合作的方式,公共资金通常与具有商业目的的私人资本混合共同开展科研项目,为避免科研机构侵权风险的发生,《指令》又明确认同了私人机构可与科研主体共同参与 TDM^[10]。这一方面导致“科研机构”这一主体范围变得模糊,另一方面还会对 TDM 行为方式产生实质性影响,似乎法律认可公私合作方式下非科研机构可以从事专属于科研机构的数据“复制和撷取”行为。其二,适用规则不明。例如,TDM 以科研为目的开展研发进而增进社会福祉是值得肯认的,但是否应当将商业利益为主要目的的私人机构纳入 TDM 著作权例外,各国存在不同做法。德国仅允许纯科学研究目的的 TDM 著作权例外,即只有推动科技文化发展的

科研主体才能成为适格主体。然而,日本为了推动本国人工智能产业的快速发展,摆脱自身在TDM领域发展整体落后的局面,不仅未对TDM著作权例外的行为主体作特别限制,还将行为目的解释为“信息分析”这一宽泛的概念,商业性目的与非商业性目的都被纳入TDM著作权例外范围。

此外,贸然设置TDM著作权例外有失“利益平衡”考量。从知识产权法“利益平衡”理论的角度去看,不论是立法上的权利义务合理配置还是法律解释和适用过程,都体现着知识产权人的专有权利和社会公众权利、个人利益与社会公共利益等社会多元利益之间的平衡关系^[11]。著作权法以赋予创作者财产权与人身权作为激励手段鼓励文化创新,秉持着“先授权再使用”的原则,例外应当建立在周延的价值判断基础上。我国若未充分考量自身国情及相应的立法价值,忽视各主体之间的复杂利益关系,贸然以著作权例外的方式将TDM行为合法化,必然会破坏创新成果保护机制,打破著作权制度中已有的平衡关系。

2.2.2 引入转换性使用缺乏法律移植土壤

首先,从立法起源上看,美国作为转换性使用的发源地,其合理使用判断一直备受争议,事实上从未形成统一判断标准。1976年美国著作权法在总结近百年判例的基础上将“四要素”判断法写入其中,然而,这种原则性的规定并未使合理使用的判断变得轻松,相反“四要素”在界定上的模糊给学界和实务界带来了诸多困惑。在著名的Sony案中,美国联邦最高法院的大法官们认为,对于私人录制行为是否构成合理使用重点在于“使用行为的目的与特性”(第一项要素)与“对作品潜在市场所造成的影响”(第四项要素)的判断^①。波斯纳大法官则认为,“使用行为的目的与特性”(第一项要素)与“作品的性质”(第二项要素)是空洞的,只有“使用作品的

比重”(第三项要素)是符合经济学规律的判断^[12]。在著作权领域影响颇深的“谷歌数字图书馆案”中,谷歌图书馆计划的大量复制行为明显不符合“适当引用”(第三项要素)的要求,且会对著作权人的潜在市场造成影响(第四项要素),但法院坚持认为对“使用行为的目的与特性”(第一项要素)的考量更为重要,商业性目的不能妨碍构成合理使用,谷歌提供作品片段的行为是对原告作品的转换性使用^②。美国学者Nimmer在总结众多相关判例后认为:“合理使用的判断标准一直未形成统一,四要素在任何一种观点中都可以得到合理主张”^[13]。可见,虽然转换性使用能够对合理使用的价值判断进行调适,使其为作品使用的新型方式提供合法性基础,但转换性使用自身的外延模糊会造成合理使用的滥用。即使是经历二十余年相关司法实践的的美国,也未能形成统一的适用标准,法官们依然面临着如何解释的难题^[14]。

其次,我国奉行严格的知识产权法定主义,主张知识产权的种类、权利客体、权利保护的范、权利的具体内容等必须通过制定法做出严格而明确的规定^[15]。如果引入合理使用“一般条款”作为TDM著作权侵权判断方式,将极大地拓展法官自由裁量空间,这就意味着法官可以解释、创设新的权利限制,TDM领域可能会成为数字时代著作权人合法利益被侵害的重灾区,行为人可以借由TDM形式侵犯他人著作权。因为法律制度移植不能脱离其适用环境,在判例法国家,先例确定下的规则可以约束其后司法裁判活动,而在我国这样的成文法国家,法院不同的解释路径会造成TDM著作权例外规则适用的不统一,引发司法裁判的冲突。

再次,从我国著作权法修订的实际看,“一般条款”并未被纳入合理使用。在司法实践中,出现过法院超越当时《著作权法》第二十二条所列的事项,径行运用美国合理使用“四要素”判断

① Sony Corp. of America. v. University City Studios, Inc., 464 U. S. 417(1984)。

② Authors Guild, Inc. v. Google, Inc. 804 F. 3d 202(2nd Cir. 2015)。

法,从而认定作品使用行为构成转换性使用的案件^①。就此有学者主张引入合理使用“四要素”判断法,在《著作权法》第三次修订的第二稿、第三稿及其后的送审稿中也都曾出现有关合理使用的“一般条款”^[16]。然而,2020年通过的最新版《著作权法》最终未采纳开放式合理使用的立法模式,以谨慎的立法态度防止合理使用泛化为“一般条款”。综上,在现阶段继续探讨增设TDM著作权例外、探讨是否引入“一般条款”似乎已不具有现实意义,本文也无意涉足广义上的TDM合法性问题,通过上述分析意在说明,如何在我国现有的著作权法框架下寻求判定图书馆TDM行为合法性的进路,对图书馆行业来说才是当务之急。

3 TDM行为本质属性分析与著作权侵权风险的发生

TDM一般包含以下四个步骤:步骤1文本与数据收集、步骤2数据预处理、步骤3建立数据模型、步骤4产生信息分析结果。具体来说,首先,文本与数据收集是通过计算机技术或其它数据收集手段提取有价值的信息,为信息分析提供基础资料。其次,数据预处理是通过数据标注、清洗降低文本中的噪声和非规范性表达,将非结构化的语言转化为计算机可以理解的结构化语言,从而改进收集数据的质量、提高数据挖掘过程的精度和性能^[17]。再次,建立数据模型是根据不同文本数据的特征建立与之相符合的主题模型,用统一、规范的数据仓库来管理数据。最后,通过数据模型分析产生情感与观点的挖掘、话题检测与追踪等信息分析结果^[1](8-9)]。

在步骤1文本与数据收集集中,TDM首先需要收集和整理信息,将知识和信息转化为计算机能够读取的数据格式。在这些海量数据中很有可能涉及到著作权法所保护的作品,如将纸质书目转化为电子文档,只要这些书目仍处在作品的

保护期内,这种“转化”就构成了著作权法意义上的“复制”^[18]。依照著作权法中作品使用必须先授权的原则,TDM主体未经许可复制他人作品则会产生著作权侵权的风险,除非图书馆仅用于自身陈列或保存文献的需要。

在步骤2数据预处理、步骤3建立数据模型中,结构化行为的实质是剔除机器分析所不需要的成分,以一种模块化的方式呈现,以下通过沈从文《边城》段落简要予以说明。

月光如银子,无处不可照及,山上篁竹在月光下皆成为黑色(主题1:时间)。身边草丛中虫声繁密如落雨。间或不知道从什么地方(主题2:地点),忽然会有一只草莺“落落落落嘘!”(主题3:动物)啾着它的喉咙,不久之间,这小鸟儿又好像明白这是半夜,不应当那么吵闹,便仍然闭着那小小眼儿安睡了(主题4:事件)。

通过数据预处理后,呈现给机器的《边城》已不再是一部作品,而是时间、地点、动物、事件这样的自然语言主题模型。这两个阶段对于作品的结构化分析,如同人对于作品的阅读与欣赏,其本身不再是著作权法意义上的作品利用形式,可被直接归于合理使用的范畴,因而无须考虑该环节的侵权风险问题^[19]。

在步骤4产生信息分析结果中,TDM所产生的信息分析结果,是一种统计学意义上的大概率事件,很多情况下计算机专家也不清楚TDM的结果是什么。在进入人工智能时代后,信息分析结果变得更加难以预测,除了极少数被严格控制的监督学习以外,我们根本不能考察静态的源代码或原始数据,无法推断机器学习算法的运算结果^[20]。因此本文主张,对于这一阶段TDM著作权侵权风险的分析与判断,应当坚持实用主义为导向,就信息分析结果的合法性进行类型化分析,旨在为TDM主体提供更为确定的侵权判断方法和更为清晰明确的侵权风险判断标准,使其避免侵权行为的发生。

^① 北京市第一中级人民法院民事判决书(2011)一中民初字第1321号。

国内邮发代号：2-997

国外邮发代号：Q1314

广告许可证：京海工商广字第 0007 号

定价：20.00 元

ISSN 1009-3125



9 771009 312128