

支持文本与数据挖掘的著作权法律政策建议^{*}

罗 娇 张晓林

摘 要 文本与数据挖掘(TDM)实现了从人工阅读到机器阅读的变革,为创新提供了新的方法和工具,因而在科研界有强烈的应用需求,但也因著作权法方面的不确定性而导致发展受限,实践界试图通过政策声明、司法个案和立法程序来解决 TDM 所面临的著作权问题。在厘清 TDM 著作权问题的基础上,本研究结合实践中的应对措施与我国的法律环境,建议通过设立 TDM 著作权例外、推动“转化性使用”理论的司法适用、借助订购协议强化权利、发展开放获取事业等多种措施来应对 TDM 面临的著作权问题,为 TDM 技术的发展与应用提供法律政策支持。图 2。参考文献 26。

关键词 文本挖掘 数据挖掘 著作权 法律政策

分类号 G254

Copyright Legislative Proposal and Policy Advice for Text and Data Mining

LUO Jiao & ZHANG Xiaolin

ABSTRACT

Text and Data Mining (TDM) generally refers to mining patterns, trends, and other valuable information from text and data by using automated analysis techniques. It is a computer-based process, which derives or organizes information from text or data. TDM is of great value for improving scientific research efficiency, accelerating scientific discoveries, promoting scientific innovations, and realizing economic growth, but it is seriously hampered by market failures, legal uncertainties, and isolated information islands. Copyright uncertainty is the most significant legal obstacle faced by TDM. However, the current domestic researches on this issue focus on tracking the development of foreign countries, and relatively lack localization research on law of China. Therefore, this paper analyzes whether the TDM infringes copyright on Chinese copyright law, and believes that the TDM involves copyright-controlled behaviors under certain circumstances, which the statutory infringement exemptions such as copyright exceptions and contract authorizations cannot cover completely. At the same time, combined with the international responses to TDM copyright legal issues and the Copyright Law of China, this paper proposes a legislative proposal and several policy advices.

For the legislative solution, it is recommended to add an TDM exception in the third revision of the Copyright Law of China, and to embody the following content in the clause: 1) it would only apply where

^{*} 本文系国家社会科学基金青年项目“图书馆数字化服务的著作权问题及其解决方案研究”(编号:17CTQ037)的研究成果之一。(This article is an outcome of the youth project “Research on the Copyright Issues and Its Solution in Library Digital Service” (No.17CTQ037) supported by National Social Science Foundation of China.)

通信作者:罗娇,Email:luoj@mail.las.ac.cn, ORCID:0000-0003-4949-5211 (Correspondence should be addressed to LUO Jiao, Email: luoj@mail.las.ac.cn, ORCID: 0000-0003-4949-5211)

justified by non-commercial purposes ; 2) it would only benefit users having a lawful access to the data; 3) it would not apply if the analysis output substitutes for the pre-existing works or databases; 4) mentioning the sources of the name of authors would not be an obligation; 5) it could not be overridden by contractual terms and technical measures; and 6) it would not apply to tools designed for TDM, and would not affect the application of privacy, confidentiality and special data protections rules.

For non-legislative solutions, in order to promote the application and development of TDM, it is recommended to eliminate the copyright uncertainty by following measures: 1) it would respect and protect users who have a lawful access to the data, on the premise that TDM is not an alternative competition with the copyright holders' original market; 2) it would understand the TDM from "transformative use" paradigm that developed in the recent judicial cases of US, and would encourage TDM within the scope of fair use; 3) it would draw support from agencies and their alliances, which have excellent negotiation experience, funding capacity and contracting ability, to emphasize that the TDM within the scope of fair uses could not be overridden by contractual terms, and even to strive for more extensive TDM rights in their contracts; 4) it would promote the development of open access for a wider application of TDM technology, especially the open access with the licenses of CC-by, CC-0, ODC-by, ODC-0, etc. 2 figs. 26 refs.

KEY WORDS

Text mining. Data mining. Copyright. Legislative policy.

文本与数据挖掘 (Text and Data Mining, TDM), 一般指利用自动分析技术分析文本与数据的模式、趋势以及其他有价值的信息, 是以计算机为基础的、从文本或数据导出或组织信息的过程^[1-2]。在信息暴增的大数据时代, TDM 对提高科研效率、加速科学发现、促进科研创新、实现经济增长具有重要价值。但我国学界目前对 TDM 法律政策的研究多集中于跟踪国外的发展动态^[3-5], 相对缺乏针对中国问题的本土化研究, 因此有必要结合我国著作权法来剖析 TDM 可能面临的著作权问题, 并提出相应的法律政策建议。

1 文本与数据挖掘对科研创新具有重要价值

1.1 实现“人工阅读”到“机器阅读”的变革

TDM 实现了“人工阅读”到“机器阅读”的变革, 大大提升了阅读效率。经济合作与发展组织 (Organization for Economic Cooperation and Development) 的数据显示, 2008 到 2011 年间产

生的数字化信息超过了以往所有的历史记录^[6]; 国际科技医学出版商联盟 (International Association of Scientific, Technical & Medical Publishers, STM) 的报告显示, 全球科研界每年将产生超过 150 万篇新的学术论文^[7]。面对如此巨量的信息, 单纯依靠人力阅读如此大规模信息的效率相对低下, 但通过 TDM 技术, 将减少人类阅读时间的 80% 并且提升数据管理 50% 的效率^[8]。在经济、学术和社会活动每天都会产生巨量信息和数据的大数据环境下, 数据极具潜在的经济和社会价值, 需要诸如 TDM 之类的技术进行发掘。随着计算机技术和人工智能的进一步发展, TDM 的需求将愈加明显。

1.2 为创新提供新的方法和工具

与基于实验室研究产生的科学发现不同, TDM 是通过已有文献来加速知识发现, 并从既有知识中发现新的知识。因此, TDM 的重要价值不仅体现在加速“阅读”的过程, 更体现在据此获得新的观点, 这一过程也被称作“数据驱动科学”^[9]。例如, 在计算机领域, 服务于特定语

义分析的专用软件工具可以作为 TDM 的一部分来使用^[9];在医药领域,通过 TDM 将现有药物和新的应用进行关联,发现了新的蛋白质、基因、疾病之间的关联^[10];在化学领域,剑桥大学的化学家皮特·穆雷-卢斯特(Peter Murray-Rust)认为,通过 TDM 对化学内容的挖掘可以创造价值超过十亿欧元的新的信息,并有助于开发新的药物以及开发更好的可重复使用的化学及其相关教育产品^[11]。欧洲研究图书馆协会(Association of European Research Libraries, LIBER)曾列举多个实践案例来证明 TDM 技术对创新的驱动效应^[11]:

European Bio-Informatics Institute(欧洲生物信息研究所)——该机构管理着世界上最大的分子数据库,它采用 TDM 技术检索 80 万篇开放获取文章中基因和炎症肠病的关联,在一天之内便发现了 6 000 个相关的例子;

Agro-Know ——由于食源性疾病每年会造成 35 万人死亡,Agro-Know 公司通过各种渠道(如出版物、政府报告、数据集、社交媒体)聚集信息,并转化为可重用格式提供给公共和私营部门的决策者和科学家,以及需要被告知食品安全问题的消费者,从而降低食源性疾病发生的风险;

Chemical Tagger——是用于化学语义文本挖掘的开源工具,已经审查了 50 多万项专利,并从大于 99.5%的精度语境中确定溶剂。

TDM 技术对于研究人员往往是不可或缺的,几乎所有研究人员都已直接或间接地使用了 TDM 技术^[12]。因为即使在自己非常熟悉的研究领域,科研人员也需要访问大量的出版物和研究数据,这往往需要通过搜索引擎或书目数据库 TDM 技术予以实现,只是科研人员也许还没有意识到类似的检索手段就是 TDM。事实上,TDM 的价值远远超出了搜索引擎或书目数据库,未来 TDM 的发展潜力在于通过人工智能、机器学习、统计学和数据库等交叉方法,以更自动化的条件挖掘更大规模的内容、得到更精准的挖掘结果,从而发现更多知识。

2 文本与数据挖掘面临的著作权问题

TDM 潜在价值的充分实现,除了需要突破技术壁垒,还需要克服数据质量参差不齐、研究者缺乏 TDM 技能和意识,以及隐私、伦理、知识产权等非技术壁垒。因此有学者指出,有望改变生活方式的 TDM 正受到市场失灵、法律不确定性以及信息孤岛的严重阻碍^[5]。其中,著作权方面的不确定性是 TDM 面临的最显著的障碍。在著作权法背景下考察 TDM 是否侵犯著作权,往往需要先分析 TDM 行为是否涉及受著作权控制的行为;若涉及,则分析这些行为是否有侵权豁免的事由,如有合同、许可协议的授权,或者属于著作权例外(如合理使用)等(见图 1)。

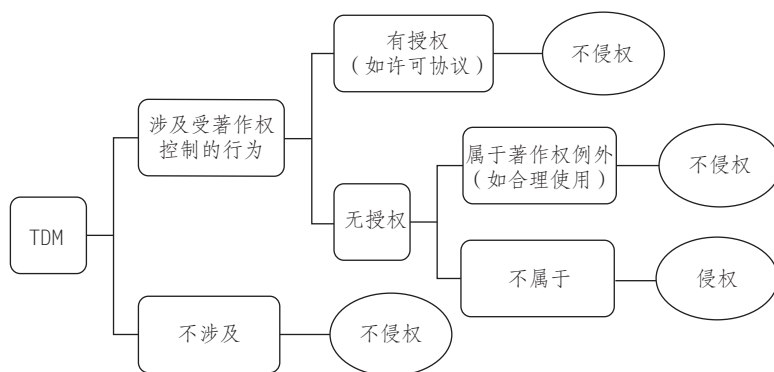


图 1 TDM 著作权侵权可能性分析

2.1 文本与数据挖掘的流程及可能涉及的排他性权利

TDM 是从大量数据中抽取特征值,并对特征值进行分析计算得到新的知识的过程^[7]。虽然不同技术下 TDM 的操作流程有所不同,但大致可概括为 4 个步骤:①“信息抽取”,即从文本中或数据集中提取有意义的信息或内容对象,并将这些对象集成于数据集(Data Set)、知识库(Repository)或集成库(Collection)中;②“语义分析”,即对抽取出来的内容对象赋予语义含义和相互关系;③“关系计算”,即计算这些对象及其关系的分布、关系、相互作用、演化等特征趋势;④“知识发现”,即根据计算得到的关系模式来发现新知识或验证研究假设^[5]。无论具体的技术方法或实施形式如何,TDM 需满足一个先决条件,即能够对被挖掘的内容进行物理访问^[6],物理访问的方式可以是将同一样本的内容进行系统性复制(并可能进行本地存档),也可以是将不同来源的内容复制到一个数据库中。

TDM 所挖掘的“文本”与“数据”,包括受著作权保护或不受著作权保护的任何形式的材料。TDM 之所以产生著作权法律争议,主要原因在于 TDM 的流程可能涉及下述受著作权控制的行为。

其一,可能涉及对受著作权保护的数据或信息的复制行为。技术上,TDM 流程往往需要将大量数据读入到系统之中,这涉及对这些内容的复制;有时,可能要将 PDF 格式的文档转换为文本格式或 XML 格式的文档才能进行数据分析,这一格式转换的过程也涉及复制行为。在传统的著作权意义上,复制行为是受著作权控制的行为之一,如果行为人既未获得著作权人的授权又没有其他侵权豁免事由,则 TDM 涉嫌侵犯著作权人的复制权。英国高等教育国际联盟因此指出,“由于计算机必须制作整个版权作品的副本才能执行相同的活动,数据挖掘的过程将受版权法的约束”^[13]。Borghi 和 Karapapa 也认为,“(TDM)自动化分析的前提是反复复制

整个作品……”^{[14]51}。

需要说明的是,并非所有的 TDM 流程都需要进行全文复制,例如软件只是“爬行”(Crawls)通过文本或数据,并“逐个”处理它们,不用复制整个文本,而是一次只复制单个或少量的单词或数据。在这种情况下,软件只“吞咽”一个或两个单词,然后继续下一个,并不保留它们的副本,而只是“计数”其出现次数,此时并不涉及著作权意义上的复制行为。

其二,可能涉及对受著作权保护的数据库的复制行为。当数据库中的数据或材料的选择、编排体现了一定程度的独创性,则数据库本身可构成汇编作品。作为汇编作品的数据库,著作权所保护的是其对数据或材料的选择、编排,而非数据或材料本身。在 TDM 流程中,如果复制整个数据库(如系统性复制),则同时也复制了数据库对数据或材料的选择、编排,后者涉及受著作权所控制的复制行为。此时,如果行为人为人既未获得数据库著作权人的授权,又没有其他侵权豁免事由,则 TDM 涉嫌侵犯相应数据库的复制权。但如果只是复制数据库的一部分,则是否构成著作权意义上的复制需结合具体情况分析,通常认为对数据库实质性部分的复制才属于对数据库作品的著作权意义上的复制。

其三,可能涉及对不受著作权保护的数据库的抽取行为。如果数据库中的数据或材料的选择、编排并不具备构成汇编作品所要求的独创性,则数据库本身不能获得著作权保护。然而,在承认数据库权的国家和地区,数据库如果存在有关数据“存储”和“处理”上的实质性投资,则可获得数据库权的特殊保护。例如,《欧盟数据库保护指令》第 7 条的规定,数据库权对不受著作权保护的数据库提供“抽取权”(the Extraction Right)和“重用权”(Re-utilization Right)两项特殊保护,前者禁止他人未经许可就将数据库内容的全部或实质性部分以任何手段或方式永久性暂时性地转移到另一媒介之中;后者禁止他人未经许可通过发行复制件、出

租、网络传输等方式,将数据库内容的全部或实质性部分提供给社会公众。在 TDM 流程中,有时会涉及“抽取”行为,如数据库内容的全部或实质性部分进行本地存档,此时若行为人既未获得数据库权利人的授权,又没有其他侵权豁免事由,则涉嫌侵犯相应数据库的抽取权。当然也不排除特殊情形下,软件只是“爬行”通过数据库,而不需“抽取”数据库内容的全部或实质性部分,则 TDM 并不涉及数据库权意义上的“抽取”行为。需要说明的是,由于我国著作权法并未规定数据库权,因此我国著作权法背景下的 TDM 行为,原则上不涉及数据库权中的“抽取”行为。

2.2 现行著作权例外不能完全覆盖文本与数据挖掘的所有行为

TDM 涉及受著作权控制的行为并不必然意味着 TDM 侵犯著作权,相反,若存在豁免的事由,如属于合理使用等著作权例外的范围,则 TDM 可具备完整的合法性依据,并不构成侵权。然而结合我国著作权法,我国的著作权例外并不能完全涵盖 TDM 的所有行为,使 TDM 面临著作权上的不确定性。

其一,现行的著作权例外难以涵盖非个人行为的 TDM。根据我国《著作权法》第 22 条第一款第一项的规定,为个人学习、研究或者欣赏,使用他人已经发表的作品,可以不经著作权人许可,不向其支付报酬。该项例外可以为基于“个人学习、研究或者欣赏”的 TDM 行为提供合法性依据,但难以为作为公司或机构的非“个人”行为的 TDM 提供合法性依据。

其二,现行的著作权例外难以涵盖非教学性或非研究性的 TDM,也难以涵盖教学性或研究性、但需大量复制作品的 TDM。根据我国《著作权法》第 22 条第一款第六项的规定,为学校课堂教学或者科学研究,可以少量复制已经发表的作品。如果 TDM 行为是基于“学校课堂教学或者科学研究”之目的,且属于对作品的少量复制,则可援引该项例外作为合法性依据。然

而实践中,且不说需要使用 TDM 技术的公司或机构往往存在“学校课堂教学或者科学研究”以外的目的,即使是具有“学校课堂教学或者科学研究”之目的 TDM,也难以满足该例外中“少量复制作品”的要求,因为 TDM 的过程往往涉及对海量内容的复制,TDM 的价值也在于通过机器自动阅读海量内容,少量内容并无 TDM 的必要。

其三,现行的著作权例外难以涵盖图书馆的 TDM。图书馆作为信息保存与传播的重要机构,各国著作权法几乎均设立了专门针对图书馆的著作权例外,以此便利图书馆的业务发展。我国《著作权法》第 22 条第一款第八项亦规定了针对图书馆的著作权例外,即为了陈列或者保存版本的需要,可复制本馆收藏的作品。同时,根据我国《信息网络传播权保护条例》第 7 条第二款的规定,为陈列或者保存版本需要以数字化形式复制的作品,应当是已经损毁或者濒临损毁、丢失或者失窃,或者其存储格式已经过时,并且在市场上无法购买或者只能以明显高于标定的价格购买的作品。TDM 涉及的复制行为显然难以满足这一要求,该项例外也难以成为图书馆 TDM 行为的合法性依据。

此外,根据我国《著作权法》第 22 条第一款的规定,适用著作权例外应当以“指明作者姓名、作品名称”为前提条件,而要求在 TDM 中一一指明其所涉及的每一部作品的作者姓名和作品名称是不现实的,因为 TDM 的分析对象往往是海量的数据。

2.3 许可协议不能完全解决问题

当 TDM 涉及受著作权控制的行为,且著作权例外并不能完全涵盖 TDM 的所有行为时,可否通过许可解决 TDM 的著作权问题?欧洲出版商协会(Europe Publishers Council,EPC)^[15]、国际科技医药出版商联盟(International Association of Scientific, Technical & Medical Publishers Council,STM)^[16]等出版商持肯定态度;国际图书馆协会联合会(International Federation of Library

Associations and Institutions, IFLA)^[2]、欧洲研究图书馆协会 (Association of European Research Libraries, LIBER)^[17] 等图书馆联盟则认为 TDM 本身具有合法性依据, 不需要进行额外的协议授权。事实上, 即使需要额外授权, 由于谈判地位不对等、谈判力量悬殊和谈判成本的限制, 许可协议也不能完全解决 TDM 的著作权问题, 甚至还会对 TDM 的应用形成限制。

其一, 许可协议可能限制 TDM 的工具和方法。部分出版商的文本与数据挖掘协议规定, 订阅用户进行 TDM 时, 只能通过其应用数据接口 (API) 获取所挖掘的内容, 并明确禁止使用机器人 (Robots)、蜘蛛 (Spiders)、网络爬虫 (Crawlers) 或者其他自动程序或算法从它的网站下载内容, 而这些都是内容挖掘最常用的方法^[17]。在条件如此严苛的许可协议下, 订阅用户所获得的挖掘权利其实非常有限。

其二, 许可协议可能限制被挖掘内容的格式。例如, 数据库商在许可协议中规定所挖掘的内容只能通过 API 获取纯文本格式的全文内容。虽然通过 API 获取内容并进行挖掘是 TDM 的常见方法, 但是现在文章的内容变得更加丰富, 包括图片、图表、互动内容以及视频, 这些图片、图表也许同文本具有一样的价值, 尤其是对于从事人类学、遗传学和化学的研究者来说, 图片和图表是内容中最有价值的部分之一, API 仅允许文本获取, 对最终挖掘成果的实际价值将产生不良影响。

其三, 许可协议可能限制被挖掘内容的数量。部分出版商在许可协议中将 TDM 的数量限制在每周一万篇文章^[17], 且不说这一数字是经过严谨的科学测算得出, 还是协议制定者随意为之, 仅从大数据环境下的数据规模来看, 这一数量也是远远不够的, 直接影响了挖掘结果的准确性。此外, 许可协议还限制 TDM 输出结果最多包含 200 个字符的原文“片段”^[17], 相当于限制了对原始资料的引用量, 也可能会导致错引, 或导致对原始研究显示的不准确。

其四, 许可协议可能限制被挖掘内容的存

储和共享。例如, 在许可协议中规定, 由被挖掘内容组成的相应数据集须在使用完毕后删除, 不得将相应数据集存储在其他机构知识库中并分享给其他研究者^[17]。这将会给其他研究者验证挖掘成果带来障碍, 挖掘结果的透明性和重用也因此无法保证。

此外, 许可协议不仅不能完全解决 TDM 的著作权问题, 还可能不必要地大量占用科研人员的科研时间。有研究显示, 如果从每一个握有版权的出版商那里寻求某一项目所需要挖掘的数千篇文章的授权, 将占用科研人员 62% 的研究时间^[5]。被挖掘的文章数量越多, 寻求许可所花费的时间也会随之增加。即使可以通过出版商或数据集成商获得批量授权, 也会面临因许可协议条款不明确和相应法律责任难以落实而带来的不可避免的时间成本。

3 实践中的应对方案

3.1 试图重新认识知识利用的权利

近年来, 国际图书馆界以发布政策声明、推动国际宣言的签署等措施, 试图重新认识包括 TDM 在内的知识利用的权利。

3.1.1 提倡有阅读权就有挖掘权

2013 年, 国际图书馆协会联合会 (IFLA) 发布文本与数据挖掘的政策声明, 提倡有阅读权就有挖掘权。IFLA 认为, 不能僵硬地理解著作权法、数据库相关法律或者与之相反的合同条款, 只要 TDM 的结果并不能替换受著作权保护的原始作品, 研究人员就能够共享这些成果; 如果研究人员或研究机构, 或者是其他通过图书馆获取信息的用户, 已经合法地获取到数据内容 (包括数据库), 那么读取的权利就应该包括挖掘的权利^[2]。欧洲研究图书馆协会 (LIBER) 也持相同观点, 提倡有权阅读就有权挖掘^[17]。英国知识产权局在阐述其 TDM 例外条款的立法理由时也指出, “如果研究人员已经有权利阅读某作品 (即他们合法获取了该作品), 则他们可以基于计算分析的目的对该作品进行复制……”^[18]。

3.1.2 推进《海牙宣言》的签署

鉴于包括著作权法在内的知识产权法对TDM形成的不当约束,2014年12月,欧洲研究图书馆协会组织来自全球多个领域的25名专家起草《数字时代知识发现海牙宣言》(以下简称《海牙宣言》)。该宣言充分认可数字化时代TDM潜在的巨大经济价值和社会效益,主张从法律、政策、基础设施和工具、宣传培训等方面来保障TDM的发展。在法律方面,《海牙宣言》倡议相关立法应当允许用户对其合法获取的材料进行TDM,著作权法应当创建TDM的著作权例外,并禁止合同或技术措施限制法律所赋予的TDM权利;在政策方面,《海牙宣言》倡议大学、科研机构、科研资助者和商业机构在尊重作者和出版者合法权利的前提下确保其政策支持TDM,并通过激励措施激励TDM的运用;在基础设施和工具方面,《海牙宣言》倡议科研机构、大学和商业机构的知识库等基础设施应当保证对合法用于TDM的出版物进行存储、访问、获取;在宣传培训方面,《海牙宣言》倡议大学、科研机构、图书馆协会、医疗界、企业对TDM的益处进行宣传,并倡议图书馆员为科研人员提供TDM知识的培训,包括法律咨询。《海牙宣言》于2015年5月6日发布,迄今为止,超过200个组织和600个自然人签署了该宣言^[19]。

3.2 以合理使用原则支持TDM

目前,美国法院在多个司法判决中支持TDM属于合理使用。由于美国属于判例法国家,这些判决所具有的既判力意味着美国以司法个案的形式认可了TDM行为的合法性。

3.2.1 美国法院支持TDM属于合理使用的司法判决

2003年至今,美国法院在多个判决中支持TDM涉及的复制行为属于合理使用。

“作家协会诉HathiTrust”案。该案中,一批高校为了创造出帮助读者检索书籍并在语料库中进行文本挖掘的数字化检索工具,对数百万受版权保护的图书进行数字化,并允许用户对这些数字化的图书进行TDM,美国第二巡回法院判决这一行为构成合理使用^①。

“White诉West”案。该案中,West和Lexis公司将包括案件摘要在内的法律公文复制到数据库Westlaw和LexisNexis中提供全文获取并创建一个TDM工具,其检索结果包含法律文件的全文。纽约州地区法院认为,“West和Lexis对文件的回顾、选择、转换、编码、连接和识别过程带来了新东西,与原始摘录相比有新目的和与之不同的特征”,因此判决其行为构成合理使用^②。

“Fox诉TVEyes”案。该案中,TVEyes录制了电视和无线电广播的全部节目内容,并运用隐藏字幕和“语音到文本”(Speech-to-Text)的转换技术,对这些内容创建了一个可检索的数据库,纽约州地区法院认为TVEyes的行为构成合理使用^③。

“作家协会诉谷歌”案。Google与一些著名高校的图书馆合作,将图书馆中的馆藏书籍进行数字化扫描并纳入一个供学者和研究人员使用的可检索的数据库中,检索结果以文本“片断”的形式展示,美国第二巡回法院判决这一行为构成合理使用^④。

① 参见美国联邦第二巡回法院“作家协会诉HathiTrust”一案判决书, *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014)。

② 参见美国纽约州南部法院“White诉West出版公司”一案判决书, *White v. West Pub. Corp.*, 2014 WL 3057885 12-civ-1340-JSR (S.D.N.Y. 2014)。

③ 参见美国纽约州南部法院“Fox News Network公司诉TVEyes”一案, *Fox News Network LLC, v. TVEyes*, 13 Civ. 5315 (AKH) (S.D.N.Y. 2014)。

④ 参见美国纽约州南部法院“作家协会诉谷歌公司”一案判决书, *Authors Guild v. Google*, 770 F.Supp.2d 666 (S.D.N.Y. 2011)。

“A.V.诉 iParadigms”案。该案中,iParadigms 创建了一个名为 TurnItIn 的数据库,该数据库允许教师将学生作品与互联网上的内容进行对比,从而判断该作品是否存在抄袭。尽管 TurnItIn 是商业性质的,但是该使用行为被认为具有“高度的转换性”而被美国第四巡回法院判决构成合理使用^①。

“Perfect 10 诉亚马逊”案。该案中,Google 在其搜索引擎中使用了原告享有版权的图像的“缩略图”,并将其“链接”到原图片,从而指引用户到原告网站上访问正常尺寸的全图。美国第九巡回上诉法院认为缩略图作为“电子索引”,对版权作品的使用具有高度转换性,构成合理使用^②。

“Field 诉谷歌”案。该案中,谷歌在它的网站缓存中提供了版权人原始的网络内容的副本,缓存的链接被用于多种用途,如存档、搜索查询中的网络分析或识别,美国第二巡回上诉法院认为该行为构成合理使用^③。

“Kelly 诉 Arriba Soft”案。该案中,搜索引擎公司 Arriba Soft 将摄影网站上受版权保护的图片复制为缩略图并设置了链接,美国第九巡回上诉法院认为该行为构成合理使用^④。

3.2.2 美国法院认定 TDM 属于合理使用的法律推理

上述案例中,美国法院分别从《美国版权法》第 107 条列举的合理使用行为的四个判断因素进行分析,尤其借助了转换性使用理论,推理出 TDM 属于合理使用^[20]。美国法院首先对 TDM 行为进行了合理使用的“四因素”分析,即

对使用行为的目的和性质、被使用作品的性质、使用的数量和质量,以及使用行为对作品的潜在市场影响进行分析。

第一,使用的目的和性质。美国法院认为,使用 TDM 技术的搜索数据库或搜索引擎,其对版权作品的复制属于转换性使用,因为“全文可检索数据库中的单词搜索的结果因其来源的页面(和图书)不同,其目的、性质、表达、含义和信息也有所不同。”^⑤

第二,被使用的作品的性质。分析作品属于纪实作品还是创作作品,对前者的使用更倾向于被认定为合理使用。上述 TDM 案例中,被告的使用行为往往涉及创作作品,这一因素似乎有利于原告,但美国法院并没有过多考虑这一因素,甚至在其他有关合理使用的判决中,这一因素也很少起到显著作用。

第三,使用的数量和质量。通常认为,合理使用是少量的而非大量使用或者包含原作品最重要部分的使用。然而,TDM 需要全文复制整个作品并将其导入数据库;如果没有对完整作品的复制,研究人员将会因无法分析作品潜在的必要部分而不能有效利用 TDM。因此美国法院认为,为了实现 TDM,在数据库中对全文进行逐字复制是合理的。

第四,使用行为对作品潜在市场的影响。即考虑使用行为是否给市场带来一个有竞争力的原作品的替代品,或其演绎作品,影响了潜在购买者购买原作品,从而减少版权人的潜在收入。美国法院认为,TDM 的高度转换性表明其

① 参见美国联邦第四巡回法院“A.V.诉 iParadigms 公司”一案判决书,A.V. v. iParadigms, LLC, 562 F.3d 630, 634 (4th Cir. 2009)。

② 参见美国联邦第九巡回法院“Perfect 10 诉亚马逊”一案判决书,Perfect 10 vs. Amazon, 508 F.3d 1146 (9th Cir. 2007)。

③ 参见美国内华达州地区法院“Field 诉 Google”一案判决书,Field vs. Google, 412 F.Supp.2d 1106 (D. Nv. 2006)。

④ 参见美国联邦第九巡回法院“Kelly 诉 Arriba Soft”一案判决书,Kelly vs. Arriba Soft, 336 F. 3d 811(9th Cir. 2003)。

⑤ 参见美国纽约州南部法院“作家协会诉谷歌公司”一案判决书,Authors Guild v. Google, 770 F.Supp.2d 666 (S.D.N.Y. 2011)。

不会给作品原有的市场带来负面影响,因为该使用不可能取代版权作品;即使使用是商业性质的,依靠“转换性”特征也足以支持 TDM 构成合理使用^①。

在对 TDM 是否符合合理使用“四因素”的分析中,美国法院发现 TDM 的复制规模与商业性,有悖于对“四因素”的传统理解。因此,美国法院逐渐从“转换性使用”理论的角度来论证 TDM 构成合理使用,要点如下。

首先,因新的、有利的目的(即转换性目的)而复制,且复制的规模对于实现该转化性目的是确实必要的,则系统的、大规模的、全文的复制可以被认为是合理的。Lexis 和 West 从美国上诉法院获取了数以百万计的法律文件,HathiTrust 数字化了百万计的图书,谷歌数字图书馆扫描了各大著名高校图书馆数以千万计的图书,这些案例中系统性、大规模、全文性的复制,似乎并不符合“四因素”所要求的使用需为“少量的、非实质性”的,但美国法院借助“转换性使用”理论突破了对“四因素”的传统理解,认为基于 TDM 目的,这些案例中的大量复制、全文复制也构成合理使用,美国第二巡回法院甚至在判决中指出“出于某种目的,可能需要复制整个版权作品,在这种情况下,因素三无法衡量对合理使用的认定”^②。

其次,复制的目的与原作品的使用目的越相异,就越不会构成对原作品的替代,即使复制具有某种程度上的商业性质。自上世纪 80 年代

的“索尼”案开始,由于美国联邦最高法院法官在附带意见中主张“对受版权保护材料的一切商业性使用都假定为不合理”^③,后续的“Harper & Row”案^④、“Acuff-Rose 音乐公司诉 Campbell”案^⑤等合理案例都坚持了这一观点。在这些案例的影响下,任何被认为是“商业”的用途几乎都不可能成为合理使用^[21]。但在 Campbell 案中,美国联邦最高法院纠正了自“索尼”案以来确立的这一错误观点,认为“转换性使用”可以突破“商业目的”的制约^⑥。在 Campbell 案的影响下,美国法院借助“转换性使用”理论,将具有转换性目的且对原作品原有市场不构成显著竞争性替代的商业性质 TDM 行为,认定为合理使用,这对激活市场活力、鼓励诸如谷歌等商业公司进行昂贵且有风险的 TDM 投资具有重要意义。

3.3 以立法纳入文本与数据挖掘著作权例外

日本和英国是通过立法将 TDM 纳入著作权例外,从而消除 TDM 的著作权障碍、赋予 TDM 完整的合法性的典型国家。

3.3.1 日本的 TDM 著作权例外

2009 年,日本在其著作权法修正案第五小节第四十七条之七中,引入“为了分析信息进行的复制等”著作权例外(以下简称“信息分析例外”),该条通常被视为日本著作权法上的 TDM 著作权例外。根据该条规定,为了利用计算机分析信息(指从多数作品或者其他大量信息中

① 参见美国纽约州南部法院“White 诉 West 出版公司”一案判决书, *White v. West Pub. Corp.*, 2014 WL 3057885 12-civ-1340-JSR (S.D.N.Y. 2014)

② 参见美国纽约州南部法院“作家协会诉谷歌公司”一案判决书, *Authors Guild v. Google*, 770 F.Supp.2d 666 (S.D.N.Y. 2011)。

③ 参见美国联邦最高法院“Sony 诉 Universal City Studios”一案判决书, *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 451 (1984)。

④ 参见美国联邦最高法院“Harper & Row 出版公司诉 Nation Enters”一案判决书, *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 562 (1985)。

⑤ 参见美国联邦最高法院“Acuff-Rose Music 公司诉 Campbell”一案判决书, *Acuff-Rose Music, Inc. v. Campbell*, 972 F.2d 1429, 1436-1437 (6th Cir. 1992)。

⑥ 参见美国联邦最高法院“Campbell 诉 Acuff-Rose Music 公司”一案判决书, *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579, 591 (1994)。

筛选出构成该信息的语言、声音、影像或者其他要素,并且进行比较、分类或其他统计分析),在必要限度内,可以将作品录入储存媒介或者进行改编(包括按照该作品创作出的衍生作品的储存),但是为了供信息分析者使用制作的数据库作品,不在此限^[22]。

日本的“信息分析例外”为 TDM 行为提供了完整的合法性依据,在消除 TDM 行为著作权方面的不确定性方面具有重要价值,具有一定的进步性。其一,该例外的适用对象并不局限于文本,而是囊括了语言、声音、影像或者其他要素构成的信息,但不适用于为了供信息分析者使用而制作的数据库作品。其二,该例外并不局限于某一种特定的数据分析方法,而是涵盖了比较、分类或其他统计分析等多种数据分析方法。其三,该例外针对的权利并不局限于复制权,还包括改编权,这意味着根据该作品创作出的衍生作品的复制,也在该例外的覆盖范围内。不过,日本的“信息分析例外”也具有一定的局限性。例如,该例外将 TDM 局限于“利用计算机”分析信息,忽视了未来的 TDM 可能会出现计算机之外的其他更先进的硬件设施。换言之,日本的 TDM 著作权例外,其弱点在于过于局限于现有技术,一旦技术发生变化,这一例外将会显得过时。

3.3.2 英国的 TDM 著作权例外

2010 年,英国《哈格里夫斯报告》建议英国政府出台一个著作权例外来允许非商业用途的 TDM。2012 年,英国知识产权局对引进该例外的影响开展了评估,认为从长远来看 TDM 会带来“社会创新和长期的经济收益”。据此,2014 年 6 月,英国《1988 年著作权、外观设计和专利法(CDPA)》第 29A 条引入 TDM 著作权例外。根据该条规定,“拥有作品合法访问途径的人对著作权作品的复制在以下前提下不构成侵权:(a)对著作权作品的复制并对记录中的任何内容进行分析,且唯一目的是非商业用途的研究;(b)副本要有充分的贡献承认(除非因实际原因或其他原因不可行)”^[23]。

英国的 TDM 著作权例外允许出于非商业研究目的对合法获取的版权作品进行文本挖掘,无论被挖掘的版权作品是否出版,也无论被挖掘的作品属于何种类型。这意味着科研人员无需额外许可,便可对其合法获取的来自不同出版商、机构知识库或其他数据库中的版权作品进行 TDM,只要相应的 TDM 属于非商业研究目的。任何合同条款与技术措施均不得排除适用 TDM 例外,否则,科研人员可通过 CDPA 第 296ZE 条寻求救济,即将投诉信寄至国务卿,由国务卿对投诉进行调查。

当然,英国的 TDM 著作权例外并非为毫无限制的绝对自由。首先,该例外仅适用于非商业研究目的,这意味着任何商业性的 TDM 涉及的复制行为必须获得著作权人的许可,否则相应复制行为将构成著作权侵权。其次,该例外仅适用于合法获取的数据或内容,这意味着只有科研人员本人或者他们所在的机构已经获得授权或已经订阅相关资源的前提下,才能对 TDM 复制的版权作品进行共享,否则分享行为也将构成著作权侵权。再次,科研人员不得规避出版商或内容提供商的合理的技术措施,否则将面临民事诉讼甚至是刑事诉讼风险,因为“出版商和内容提供商可采取合理措施来维护其网络的安全性或稳定性”^[24]。第四,根据管辖原则,英国国籍之外的研究者通过英国机构可合法获取被挖掘内容的,其依据该例外进行的复制只能由其英国同事在英国完成。最后,TDM 著作权例外条款要求充分承认被复制作品的贡献,除非贡献承认难以实现。由于 TDM 往往需要组合成百上千贡献者的作品,对所有来源都进行贡献承认难以实现,实践中可通过注明其使用的数据库或者数据源的方式进行贡献承认^[25]。

除了日本和英国,欧盟也在 2016 年发布的《数字单一市场版权指令修改草案》中引入了 TDM 著作权例外。

4 我国的应对建议

正如英国政府在其著作权政策《数字机遇》

(Digital Opportunity)中指出的那样,“技术提供了代替人工阅读所有文献的服务”^[26],人们已经越来越认识到 TDM 在提高研究效率、发掘隐藏信息、开发新知识、探索新视野、改进研究和证据基础、改进研究过程和质量等方面的价值。我国应当有效借鉴实践中的应对方案,消除 TDM 面临的著作权不确定性,为应用与发展 TDM 技术提供法律和政策支持。

4.1 立法建议

与通过政策声明或联合宣言等不具有强制执行力的“软法”来争取 TDM 权利不同,通过立法设立的 TDM 著作权例外具有强制执行力,可在相当程度上消除 TDM 面临的著作权不确定性。建议我国在著作权法的第三次修改中增设 TDM 的著作权例外。

著作权法从来不是用来限制作品的阅读方式或遏制新思想、新知识的工具,也没有任何证据可以证明 TDM 会替代原作品的市场。TDM 技术广受科研人员欢迎却屡次遭遇出版商提出著作权方面的质疑,是因为出版商掌握着作品传播的技术与渠道,著作财产权一般集中在其手中;为了维持市场竞争优势,出版商往往对一切可能导致其对作品传播的范围和方式失去控制的新兴技术持警惕态度。然而,出版商和研究人员并非对立的关系,出版商需要研究人员,研究人员也需要出版商。通过立法设立 TDM 著作权例外,并不是要侵吞包括出版商在内的著作权人的既有权利,而是要谨慎地建立起著作权人与作品受众之间的平衡关系。因此,设立 TDM 著作权例外,需评估该例外对不同的利益相关者可能产生的影响,使具体的法律条款内容能兼顾不同利益相关者的利益诉求。例如,对于出版商而言,针对数据资产(如期刊、数据库)颁发 TDM 许可是其潜在的收入来源,其更加关心设立 TDM 著作权例外对其经济收益可能产生的负面影响;而对于作者,TDM 只是对其作品的偶然使用,可能其并不关心是否通过颁发 TDM 许可获得许可费。结合日本、英国 TDM 著

作权例外的立法经验,并在遵守《伯尔尼公约》《与贸易有关的知识产权协议》《世界知识产权组织版权条约》的“三步检验法”相关要求,即“TDM 著作权例外立法是在特定情况下做出、与作品的正常利用不相冲突,以及没有无理损害权利人合法利益”的前提下,建议我国在著作权法的第三次修改中增设 TDM 著作权例外,并将以下理念融入 TDM 著作权例外条款内容的设计中:其一,TDM 著作权例外适用于非营利目的的 TDM(即通过“三步检验法”);其二,TDM 著作权例外以合法获取内容作为例外的适用要件;其三,TDM 著作权例外适用于分析结果未对内容的原始来源形成竞争性替代的情形;其四,TDM 著作权例外不强制要求指明作者姓名和作品名称;其五,合同和技术措施不得排除或限制 TDM 著作权例外的适用;最后,TDM 著作权例外不适用于数据分析的工具,不影响隐私、保密和特殊数据保护规则的适用。

4.2 政策建议

鉴于设立 TDM 著作权例外需经历的立法程序复杂漫长,而 TDM 技术的发展却日新月异,TDM 应用的需求与日俱增,且不说经过漫长等待后的立法是否能够成功,即便最终立法成功,也不能保证基于以往经验的立法能与面向未来的技术发展完全契合。然而,诸如 TDM 之类有益于社会福祉的新技术,不应桎梏于法律的僵硬性与滞后性。因此,建议通过多项政策,逐渐消除当前 TDM 面临的著作权问题,进而推动 TDM 技术的应用与发展。

其一,支持以新兴技术进行的知识利用。从其他国家目前已发生的与 TDM 有关的著作权争议来看,对 TDM 行为提出著作权侵权质疑的,多集中在出版商方面,而出版商反对的原因多源于其试图占有新技术条件下因作品利用方式革新所产生的新收益。从 TDM 各方利益诉求来看(见图 2),出版商追求的是经济效益的最大化。为了维持市场竞争优势,防止对作品传播范围和方式失去控制,其往往以著作权来阻止运

用新的知识利用方式。事实上,只要不直接代替原作品或破坏原作品的市场,人们对自己合法获得的作品的利用理应受到尊重和保护。TDM 作为新技术条件下发展出的一种新型的知识利用方式,在尚未与原作品市场形成替代性竞争的前提下,亦应受到尊重和保护,因为著作权的保护范围不能扩张到人们对知识的正当利用之上。

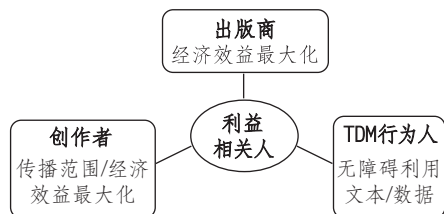


图2 TDM 利益关系

其二,鼓励符合“合理使用”制度立法精神的TDM行为。如前所述,当TDM行为涉及受著作权控制的行为时,符合著作权合理使用制度精神的TDM行为应当得到鼓励。我国现行著作权法对著作权合理使用的范围采取“列举式”的立法技术,但对未列举的可能与合理使用制度精神相符的其他行为,又缺乏灵活的认定因素。如此“非列举即排除”的合理使用条款,由于适用范围过于僵硬、难以满足新技术条件下作品利用方式变化的需求而广受诟病。虽然我国法院在其司法政策中引入了美国版权法上的合理使用判断“四因素”规则(最高人民法院《关于充分发挥知识产权审判职能作用推动社会主义文化大发展大繁荣和促进经济自主协调发展若干问题的意见》(法发〔2011〕18号)第8条),然而,TDM行为因涉及系统的、大规模的、全文的复制,或因具有某种程度上的商业因素,而难以通过传统的“四因素”规则认定其是否构成合理使用行为。因此,建议借鉴美国在TDM著作权新近司法实践中发展起来的“转换性理论”范式,从合理使用制度立法精神的角度,即促进作品的传播与利用的角度,肯定TDM所具备的“转换性使用”目的及其与合理使用制度立法精神的契合,鼓励在合理使用范围内开展TDM行为。

其三,借助资源订阅协议强化支持“合理使用”范围内的TDM,甚至争取更少限制的TDM。由于著作权法及其合理使用条款均具有一定的模糊性,出版商可以在相关合同或政策中对人们利用作品的方式进行约束,TDM因此也有受到出版商合同或政策不当限制的潜在风险。当TDM受到合同或政策的不当限制时,以个人的力量寻求通过许可的方式来获得出版商或数据库商的TDM授权,不仅会因为谈判地位不对等、谈判能力悬殊、对话机制缺乏而承受谈判失败的风险,而且即使谈判成功,仍会面临挖掘方法、工具、结果利用等被许可协议不当限制的问题。因此,建议文献资源订阅机构及其联盟借助自身作为资源订购方所具备的谈判地位、谈判经验、经费能力和签约能力,在机构及其联盟的资源订购协议中强调符合著作权“合理使用”精神的TDM等资源利用方式不受合同的限制;甚至在资源订购谈判中为用户争取超出合理使用范围的更广泛的TDM权利。

其四,通过发展开放获取事业推动更大范围的TDM应用。以立法的形式将TDM列入著作权例外,以“合理使用”制度精神鼓励TDM行为,以资源订阅协议支持TDM行为,这些应对方案虽具有一定的可行性,但也无法避免各自的局限性。首先,以立法的形式将TDM列入著作权例外,会面临成文法固有的缺陷。例如,立法漫长周期与技术的快速发展相悖,立法难免滞后;再如,立法的稳定性要求与技术的迭代性相悖,立法难以适应技术的迭代,而频繁地修法又会使立法被技术绑架;又如,立法的地域性与技术应用的全球性相悖,当一部分国家同意将TDM列入著作权例外而一部分国家拒绝这样做时,国际合作中的TDM应用仍然不能消除著作权上的不确定性。其次,以“合理使用”制度精神鼓励TDM行为,则难以判断哪些TDM行为因涉及商业性质而与合理使用制度精神相悖,更难以囊括那些具有某种商业性质的TDM行为,而后者与科研、教育性质的TDM一样,同样具有重要价值。最后,以机构的资源订阅协议争取

更广泛的 TDM 权利,且不说资源订阅机构是否具有足够的谈判能力实现这一目标;即使谈判成功,受益者也多局限于合同所列明的用户(如机构的工作人员),而难以惠及广大公众、支持公众的 TDM 应用需求。鉴于开放获取运动能够通过其运行机制,通过格式化的许可协议明确作品在不同条件下的利用权限,并力促公众对作品的自由获取,可以在相当程度上消除作品利用所面临的著作权的不确定性。因此,建议大力推动开放获取事业的发展,尤其推动以适用 CC-by、CC-0、ODC-by、ODC-0 等开放程度

较高的许可协议为代表的开放获取,通过促进作品自由获取与利用来充分推动 TDM 技术在更广泛层面的应用。

著作权法在某一时期的规定,是多方利益主体在经济、社会与技术条件下的博弈结果,但发展会不断打破这种具有时代烙印的博弈。当法律与发展产生冲突时,法律的本意与精神是保护和支持发展,而不是相反。综上,我国的著作权法律与政策应当支持诸如 TDM 之类的新兴技术条件下产生的新型知识利用方式,这对广泛应用 TDM 技术、激励 TDM 技术研发具有重要意义。

参考文献

- [1] UK Intellectual Property Office. Guidance of exceptions to copyright [EB/OL]. [2017-09-15]. <http://www.ipo.gov.uk/guidance/exceptions-to-copyright>.
- [2] IFLA. IFLA Statement of on text and data mining [EB/OL]. [2017-09-15]. https://www.ifla.org/files/assets/clm/statements/iflastatement_on_text_and_data_mining.pdf
- [3] 周玲玲. 欧盟文本与数据挖掘新策解析 [J]. 图书馆建设, 2017 (7) : 19-30. (Zhou Lingling. Study on European commission's copyright exception for researchers and text and data mining [J]. Library Development, 2017 (7) : 19-30.)
- [4] 茹丽洁, 顾立平, 田鹏伟. 国际出版商对文本和数据挖掘限制的正当性辨析 [J]. 图书馆建设, 2016 (7) : 27-33. (Ru Lijie, Gu Liping, Tian Pengwei. Analysis on properness of international publishers restricting text and data mining [J]. Library Development, 2016 (7) : 27-33.)
- [5] 徐轩, 孙益武. 英国数据挖掘著作权例外条款研究及其启示 [J]. 图书馆建设, 2015 (9) : 10-14. (Xu Xuan, Sun Yiwu. Research on UK copyright exceptions to data mining and its enlightenment [J]. Library Development, 2015 (9) : 10-14.)
- [6] World Economic Forum. Global information technology report 2012: living in a hyper-connected world [EB/OL]. [2017-09-15]. <http://reports.weforum.org/global-information-technology-2012/#section=chapter-1-4>.
- [7] Ware M, Mabe M. The STM report: an overview of scientific and scholarly journal publishing [EB/OL]. [2017-09-15]. http://www.uicc.org/sites/main/files/atoms/files/2015_02_20_STM_Report_2015.pdf.
- [8] Diane McDonald. Value and benefits of text mining [EB/OL]. [2017-09-15]. <http://www.jisc.ac.uk/publications/reports/2012/value-and-benefits-of-text-mining.aspx>.
- [9] Science Europe Working Group on Research Data. Text and data mining and the need for a science-friendly EU copyright reform: D/2015/13.324/1 [EB/OL]. [2017-09-15]. http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/WGs_docs/SE_Briefing_Paper_textand_Data_web.pdf.
- [10] UK Government. Text mining and data analytics in call for evidence responses [EB/OL]. [2017-09-15]. <http://www.ipo.gov.uk/ipreview-doc-t.pdf>.

- [11] LIBER. A copyright exception for text and data mining [EB/OL]. [2017-09-15]. <http://libereurope.eu/wp-content/uploads/2015/11/TDM-Copyright-Exception.pdf>.
- [12] Science Europe. Text and data mining and the need for a science-friendly EU copyright reform [EB/OL]. [2017-09-15]. http://www.scienceurope.org/uploads/PublicDocumentsAndSpeeches/WGs_docs/SE_Briefing_Paper_textand_Data_web.pdf.
- [13] Universities UK and UK Higher Education International Unit. European Commission's stakeholder dialogue 'Licenses for Europe' and text and data mining [EB/OL]. [2017-09-15]. <http://international.ac.uk/media/2243028/Briefing%20-%20Licenses%20for%20Europe%20and%20Text%20and%20Data%20MiningREVISED.pdf>.
- [14] Borghi M, Karapapa S. Copyright and mass digitization: a cross-jurisdictional perspective [M]. New York: Oxford University Press, 2013: 51-63.
- [15] Europe Publishers Council. Copyright enabled on the network [EB/OL]. [2017-09-15]. http://www.leru.org/files/general/EPC%20Copyright%20Vision%202014_final.pdf.
- [16] STM. Text and data mining in the EU: legal certainty already exists [EB/OL]. [2017-09-15]. http://www.stm-assoc.org/2015_05_19_New_Release_Text_Mining_in_the_EU_STM_Association_Reaction.pdf.
- [17] LIBER. Response to elsevier's text and data mining policy [EB/OL]. [2017-09-15]. <http://www.libereurope.eu/news/liber-response-to-elsevier's-text-and-data-mining-policy>.
- [18] UK Intellectual Property Office. Guidance: exceptions to copyright [EB/OL]. [2017-09-15]. <https://www.gov.uk/guidance/exceptions-to-copyright>.
- [19] The Hague Declaration [EB/OL]. [2017-09-15]. <http://thehaguedeclaration.com/>.
- [20] ARL. Text and data mining and fair use in the United States [EB/OL]. [2017-09-15]. <http://www.arl.org/storage/documents/TDM-5JUNE2015.pdf>.
- [21] Neil W N. Making sense of fair use [J]. Lewis & Clark Law Review, 2011, 15: 715-722.
- [22] Copyright Research and Information Center. Copyright law of Japan [EB/OL]. [2017-09-15]. http://www.cric.or.jp/english/clj/cl2.html#cl2_1+A47septies.
- [23] UK copyright, designs and patents act 1988 [EB/OL]. [2017-09-15]. <http://www.legislation.gov.uk/ukpga/1988/48>.
- [24] UK Intellectual Property Office. Exceptions to copyright: research [EB/OL]. [2017-09-15]. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf.
- [25] Jisc. The text and data mining copyright exception: benefits and implications for UK higher education [EB/OL]. [2017-09-15]. https://www.jisc.ac.uk/guides/text-and-data-mining-copyright-exception#footnote2_rtsudbk.
- [26] Ian Hargreaves. Digital opportunity: a review of intellectual property and growth [EB/OL]. [2017-09-15]. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf.

罗 娇 中国科学院文献情报中心助理研究员。北京 100190。

张晓林 中国科学院文献情报中心教授。北京 100190。

(收稿日期:2017-10-11;修回日期:2018-03-21)