

网络爬虫的刑事违法性边界研究

● 张良坤



[摘要] 网络爬虫技术作为一种自动化数据搜集的手段,在大数据时代得到了广泛应用。然而,随着不当使用网络爬虫行为的频繁出现,网络爬虫的危害也日益凸显。因此,有必要对网络爬虫进行法律规制,实践中也出现了众多涉及网络爬虫的司法裁判案例。然而,网络爬虫的司法适用存在一定的随意性,对于合法网络爬虫以及非法网络爬虫的界限划分并不明确。此外,网络爬虫在民事和刑事领域的界限也不够清晰。这种情况无疑会对网络爬虫的合理使用造成阻碍。因此,有必要对网络爬虫的刑事违法性边界进行深入研究。

[关键词] 网络爬虫; 刑事违法性; 数据法益

大 数据时代已经到来,无论是对于个人还是国家而言,数据都已经成为一项重要的生产要素。理论上,在现有的分配方式下,只要掌握足够多的生产要素,便能够获得相应的收益。网络爬虫技术作为一种数据搜集工具,其效率远超人工操作。通过预先设定规则,它可以自动且高效地获取所需数据,这无疑是大数据时代的一大优势。然而,在错误观念的引导下,网络爬虫技术有时被用作肆意敛财的工具,使用该技术应当遵守的道德底线和法律界限被忽视,甚至被不断侵犯,其法律危险也在不断扩大。随着网络爬虫技术的日益成熟,网络爬虫的种类和功能也在不断丰富和增强。其应用场景不再简单地局限于大型搜索引擎,而是全方位融入人们的生产生活中。只要是数据所及之处,就有网络爬虫技术施展的空间。因此,面对网络爬虫技术,既要利用它,同时也要对网络爬虫的风险进行有效地防控。

Q 网络爬虫概述

(一) 网络爬虫的概念及分类

网络爬虫作为自动收集数据的程序,其基本原理是按照使用者的需求建立需要待爬取的 URL 队列,然后从中提取出 URL,访问该 URL 所对应的网页,从网页上提出新的 URL 并放入最初的队列中。这个过程会不断循环,直到实现预定的目标。由于网络上的内容持续更新,有一类爬虫专门用于爬取网页上新产出的内容。第四类网络爬虫针对的是深层网页的数据。互联网网页按照存在方式可以分为深层网页和表层网页。表层网页可以通过静态链接进行获取,而深层网页需要提供相应的关键词或填写表单后才能访

问。至今,网络爬虫已经发展得相当成熟。

(二) 网络爬虫限制措施

数据作为一项生产要素,具有重大的经济价值。因此,对于网络爬虫的数据抓取行为,许多平台会采取多种反爬虫措施进行限制。相应地,针对这些限制,爬虫设计者也开发了许多反爬虫措施来突破此种限制。下文将对一些常见的技术措施进行介绍。

爬虫协议,又称机器人协议或爬虫规则,是由网络服务提供商依据 robots.txt 文件指导网站应对爬虫程序的协议。这一协议最初是为了应对网络爬虫所造成的网络拥堵问题而提出,其目的在于规范网络爬虫的使用行为,以避免对网站造成额外的负担。爬虫协议通常由网站所有者放置于网站根目录下,供爬虫程序识别。该协议通常由访问主体和访问范围两部分构成。关于爬虫协议的法律性质,目前尚未形成统一的意见。主流观点认为,它属于商业道德的范畴,既不具备技术上的强制力,也不具备法律上的强制效力。

相较于爬虫协议,其他的反爬虫措施都具备相应的技术反制能力。常见的反爬虫措施包括 IP 限制、验证码、登录验证、数据伪装等。当 IP 地址的访问次数和频率异常时,网站就会对该 IP 地址进行封禁,使用者就无法再通过该 IP 地址对网站内容进行访问。针对这一限制,爬虫可以通过隐藏或虚构 IP 地址来避开。验证码这种反爬虫措施在日常应用中更是广泛,通常需要使用者输入特定的数字、字母或者文字,又或者需要使用者根据提供的规则进行图形滑动、数字计算、图片选择等操作,以此来分辨是正常用户的访问行为还是爬虫的访问行为。此种限制方式可以通过人工识

别、更换 IP 地址、模拟用户等方式进行突破。除了验证码，还有登录限制这一种常见方式，其需要账号密码登录后才能访问到相应的数据或信息。这种限制只需要更换账号密码即可。因此，实践中倒卖账号密码的情形并不少见。除了对爬虫访问行为进行限制，还可以对数据进行伪装。即使爬虫可以成功访问，但无法获取到真实的数据，从而使网络爬虫的目的落空。

Q 网络爬虫的法律规定现状

(一) 网络爬虫的法律规定现状

法律的范围因观察角度的不同而有所差异。广义和狭义之分是划分法律具体涵盖范围的一个常见方法。本文所谈及的法律规定现状，是基于狭义法律的视角。目前，我国并没有专门针对网络爬虫的法律。规制网络爬虫的法律规定散见于各个部门法中。在民事领域，主要适用的法律有《民法典》《反不正当竞争法》；在行政法领域，有《个人信息保护法》《数据安全法》《网络安全法》。在民事和行政领域的法律中，许多规定是在近些年来才正式公布并实施的。在刑事领域，主要是《刑法》在网络爬虫行为及相关罪名进行规制，其中也出现了一些新的罪名，或者对传统罪名的适应性修改。总体而言，关于网络爬虫的法律规定，并没有形成一个专门的体系，而是散乱分布于各个法律部门的相应章节中。

(二) 网络爬虫的刑事司法适用现状

通过以“网络爬虫+刑事案由”为关键词进行搜索，发现相关案例仅有 10 件。为扩大搜索范围，将关键词“网络爬虫”替换为“爬虫”，刑事案件的数量增加到了 93 件。以“爬虫”与“刑事案由”为关键词进行深度检索，截至 2024 年 11 月，共发现了 93 起相关案件。经过仔细筛查并剔除了与主题无关以及重复的案件后，剩余的有效案件数量减少至 64 件。

在这些有效案件中，发现有些罪名的出现频率是远高于其他罪名，这些罪名应当是当前关注的重点。具体来说，以“侵害公民个人信息罪”定罪的案件共有 26 件，这表明在网络爬虫的相关案件中，个人信息是网络爬虫行为的重点侵害对象。同时，以“非法获取计算机信息系统数据罪”定罪的案件有 12 件，这表明计算机系统安全和以计算机系统为存储介质的数据也受到了网络爬虫的威胁。由此可知，个人信息和计算机信息系统数据成为网络爬虫行为侵害的主要对象。除了上述两个主要罪名外，其他几种类型的计算机犯罪也占据了一定比例，如“破坏计算机信息系统罪”“非法侵入计算机信息系统罪以及提供侵入”“非法控制计算机信息系统罪”定罪的案件有 5 件。这些案件表明网络爬虫的行为方式多样，对计算机信息系统的安全性和稳定

性构成了多方面的威胁。这种现象与网络爬虫技术作为计算机信息领域的一项技术有着密不可分的关系。

除此之外，网络爬虫还涉及其他罪名。以“传播淫秽物品罪”以及“传播淫秽物品牟利罪”定罪处罚的案件共有 6 件。值得注意的是，一些案件还涉及诈骗罪、盗窃罪以及开设赌场罪等较少见的罪名。另外，以“侵害著作权罪”定罪的案件虽然仅有 2 件，但也同样不容忽视，它们揭示了网络爬虫技术在侵犯知识产权方面的潜在风险。上述数据显示，围绕网络爬虫技术，已经形成了相对完整的产业链，不仅包括网络爬虫技术直接涉及的罪名，也有利用网络爬虫技术为其他犯罪提供“便利”的罪名，所以在网络爬虫行为涉及的罪名上，会存在极大的跨度。在上述案件中，有些案件存在竞合问题，即一个行为同时触犯了多个罪名。

Q 网络爬虫规制困境成因分析

(一) 网络爬虫技术专业壁垒引发的认知障碍

专业壁垒是对不同专业在深入发展之后所形成的壁垒效应的简明说法。不同的专业都有一定的专业壁垒，专业壁垒的高低也决定了其他人进入这个专业领域的难易。网络爬虫技术，作为计算机科学领域的一个重要分支，其专业性和技术难度自然形成了较高的专业壁垒。这不仅体现在对编程语言、数据结构、算法等基础计算机科学知识的掌握上，还涉及对网络协议、数据解析、信息检索等高级技术的深入理解。在当前的教育背景下，国家越来越重视人才的综合素质培养，鼓励跨学科学习和创新思维的发展。这种教育模式旨在打破传统的专业壁垒，培养能够适应多变工作环境和解决复杂问题的复合型人才。在法学专业的培养上，也呈现出这样的特点。然而，尽管有这样的教育导向，不同专业领域之间的知识差异和技能要求仍然存在，并形成了一定的隔阂。相较于传统犯罪，网络犯罪具有极强的隐蔽性和技术性，同时其犯罪成本相对低廉，但是获利可能巨大。

网络爬虫技术，作为大数据时代的关键工具，虽然起源较早，但其在刑法领域的规制却是一个相对较新的现象。司法人员在处理这类案件时，不仅要克服专业壁垒带来的理解难题，还要面对缺乏足够实务经验的挑战。随着大数据技术的快速发展，网络爬虫技术的更新迭代速度日益加快，这使得司法人员在掌握和更新相关知识方面面临更大的压力。在大数据时代，网络爬虫技术的发展不仅迅速，而且往往伴随着其他科学技术的进步，如人工智能、机器学习、云计算等。这些技术的融合使得网络爬虫的功能更加强大，应用场景更加广泛，同时也带来了更加复杂的法律问题。

法律工作者应当不断学习和适应这些快速变化的技术，

以便更好地理解和应用相关法律，公正地审理案件。然而，要实现这一目标所需的时间和经济投入是巨大的，往往超出了预期。在当前的司法实践中，司法资源的紧张状况尤为突出，案件的数量持续攀升，而办案人员的增加数量却远远跟不上案件增长的步伐。这种供需不平衡导致了司法系统的超负荷运转，影响了案件处理的效率和质量。这一现象也催生了合同制以及聘任制法律辅助人员。在资源有限的情况下，合理分配和利用司法资源也显得尤为重要。因此，为了更有效地应对网络爬虫行为所引发的法律问题，必须加强对此类行为特性及其法律后果的理解。这包括认识网络爬虫技术的多样性和复杂性，以及它们在不同案件中的不同应用。通过深入研究，可以更好地把握网络爬虫行为的法律边界，从而在司法实践中做出更加准确和合理的判断。

（二）数据法益的模糊定位

法益即法律所保护的利益，其在刑法学科体系中具有重要的地位。法益是刑法中的核心概念。法益对于成立犯罪以及罪名区分具有重要的意义。具体而言，如果一个行为没有侵害法益，其大概率不会被认定为犯罪，也就不具备成立犯罪的条件。因此，如果一个行为构成犯罪，必须确定该行为具备法益侵害性。传统法益具有相对较强的辨识性，因此传统罪名的区分也相对容易。然而，随着社会经济的发展，许多新型法益也应运而生。数据法益就是一种新兴法益，但目前对于这一法益的定位并不清晰。

首先，对于数据的概念存在一定的争议。《数据安全法》中对于数据的定义是任何以电子或其他形式对于信息的记录。对法律条文的文义进行分析，可知数据和信息是载体与内容的关系。因此，数据不等于信息，故在行政法律层面，数据和信息有其区分的必要性。然而，数据和信息的关系的密切关系在一定程度上比他们的区别更为重要。因此，有学者将数据和信息混用，甚至将其等同。

其次，对于数据法益的定位及其具体内容也存在争议。

尽管数据法益作为一种独立法益的呼声日益高涨，但对于数据法益是否具备成为独立法益的资格这一问题仍然存在争议。数据法益的具体内容也因对其定位的不同而有所不同。如果选择将数据法益作为一种单独的法益，则更偏向于赋予数据法益独特的内涵，如一些学者倡导的数据安全法益。如果否认数据法益的独立性，则会将数据法益依附于传统法益进行解释，认为数据只是传统法益的新型载体，数据本身不具备所谓的独立价值。

Q 网络爬虫的刑事违法性边界

刑事违法性是犯罪的一个重要特征。如果一个行为不

具备刑事违法性，则该行为不应认定为犯罪。违法性并不是刑事领域的专有名词，在民事领域、行政领域同样存在违法性的问题。对于同一案件，可能涉及不同法律部门的违法性。所触犯违法性的不同，可能会对案件的性质产生截然不同的影响。对案件的定性不同，则可能对案件的法律责任产生不同的影响。

最初，网络爬虫的相关案件大多是民事案件。随着网络爬虫危害性的增加，其法律风险也逐步增加，其逐步被纳入刑事法律规制的范围之中。因此，司法案例中的许多案件，其违法性的归属都比较模糊，因此存在对于网络爬虫行为扩大化打击的趋势。为此，如果需要对网络爬虫行为的案件进行准确定性。为了准确划定网络爬虫的刑事违法性边界，需要对网络爬虫行为的行为阶段进行划分。网络爬虫行为可以分为访问阶段以及数据获取阶段。在网络爬虫的行为过程中，访问阶段和数据获取阶段是连续且相互依赖的。数据获取行为必须在合法访问的基础上进行，没有合法的访问，获取数据的行为也不可能合法。因此，在评估数据获取行为的合法性时，首先必须先考察其访问权限的合法性。访问行为通常涉及两种情况：无权访问和越权访问。在无权访问的情况下，需要判断爬虫是否成功获取了数据。如果爬虫获取了数据，尤其是当数据具有明确的指向性，如公民个人信息等敏感信息时，应当根据具体罪名的构成要件进行评估，而不是简单地依据计算机信息系统数据的一般规定来判断。越权访问则是指在某些数据服务提供者允许用户访问部分数据的同时，对其他数据设置了权限限制。在这种情况下，即使爬虫拥有对部分数据的访问权限，但如果未经授权访问了限制数据，其行为可能构成违法。对于网络爬虫行为的判断标准，应当以相对可以量化的客观审查机制为准。不宜将网络爬虫行为民事违法性简单地等同于其刑事违法性的判断，而应当将其作为考量的一个方面，并对网络爬虫的刑事违法性进行独立的判断。

■ 参考文献

- [1]贺思聪.爬虫实战从技术到产品[M].北京:电子工业出版社,2019.
- [2]陈伟,孙梦蝶.基于网络爬虫技术的大数据审计方法研究[J].中国注册会计师,2018(07):76-80.
- [3]姚万勤,宋俊豪.在技术与法律之间:网络爬虫刑法规制的边界[J].法治论坛,2023(01):3-17.

作者简介:

张良坤(2000-),男,汉族,四川广安人,硕士研究生,青海民族大学法学院,研究方向:刑法学。