

数据爬取行为的刑法规制研究

熊 怡

贵州大学法学院, 贵州 贵阳

收稿日期: 2024年9月4日; 录用日期: 2024年11月8日; 发布日期: 2024年11月15日

摘 要

数据爬取行为的实质是一种从互联网获取、提炼和保存信息的自动化程序, 应当遵守技术中立原则。一旦发生技术滥用, 恶意爬取行为将会导致计算机系统信息安全以及数据所附价值产生风险。当前刑法规制存在保护法益不清、违法判断不明以及刑法前置问题。数据爬取行为的刑法规制应当明确其法益为数据承载信息的实质价值, 而非载体安全; 以整体法秩序为违法判断标准, 合理定罪量刑; 坚持刑法谦抑性原则, 建立刑法为保障法的保护体系。在规制网络爬虫行为上实现数据保护与共享的平衡。

关键词

网络爬虫, 数据安全, 刑法规制, 数据垄断

Research on Criminal Laws and Regulations of Data Crawling Behavior

Yi Xiong

Law School, Guizhou University, Guiyang Guizhou

Received: Sep. 4th, 2024; accepted: Nov. 8th, 2024; published: Nov. 15th, 2024

Abstract

Data crawling is essentially an automated process of obtaining, extracting and preserving information from the Internet, and should comply with the principle of technology neutrality. In the event of technological abuse, malicious crawling will lead to risks to the security of computer system information systems and the value attached to the data. The current criminal law system is characterized by unclear protection of legal interests, unclear judgment of violation of the law, and the problem of criminal law precedence. The criminal law system of data crawling behavior should be clear that its legal benefit is the substantive value of the information carried by the data, rather than the carrier security; determine guilt and impose appropriate sentences based on the standard of legal order;

adhere to the principle of modesty of the criminal law, and establish the criminal law for the protection of the law protection system. A balance between data protection and sharing should be achieved in regulating the behavior of network crawlers.

Keywords

Web Crawlers, Data Security, Criminal Law Regime, Data Monopolization

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在数字经济迅猛发展的当下，坐拥海量商业价值和开发潜力的信息数据可以作为一项独立的生产要素参与收入分配，数据产品作为交易客体也获得了普遍的市场认同。“数字技术发展之基在于数据。数据是生产生活与经济社会发展‘足迹’的数字记录，是数字经济时代不可或缺的生产资料和战略资源。”[1]而数据基础制度建设的根本在于安全，在全领域倡导强化数据安全的理念下，司法机关亦在采取愈发“严厉”的态度惩治破坏数据安全类型犯罪。

在云计算、区块链、人工智能等新型应用技术的加持下，数据挖掘、数据建模分析、数据预测以及决策能力已经成为高新企业提升自身业务能效的关键技能，算法甚至已经演变为一种权力[2]，超越了既有法律的规制范围。因此，刑法规制数据爬取行为的边界以及路径成为当前一大难题。

2. 数据爬取行为刑法规制证成

2.1. 网络爬虫运作逻辑结构

数据爬取行为，通常称为“网络爬虫”，是指利用计算机程序从各类数据网站、手机 APP、小程序以及搜索引擎中检索、提取、交换和收集数据的行为[3]。网络爬虫实质上就是运用计算机程序自动快速地模仿人类浏览网页并获取数据。网络爬虫程序的每一步都沿着人类通过浏览器获取数据的流程进行。

其运作机制为，网络爬虫从种子 URLs (Uniform Resource Locator) 出发，从初始网页上获取 URL，在爬取网页过程中，网络爬虫会持续从当前页面接续出新的 URL 并将它们添加到待处理队列中，这一过程不断进行，直至达到系统预设的终止条件[4]。根据功能和应用场景，网络爬虫可细分为四大类型：一是通用网络爬虫。其目标资源覆盖全网，爬取数据数量庞大且范围广泛，主要用于为门户网站搜索引擎及大型 Web 服务提供商进行数据采集[5]。二是聚焦网络爬虫。它针对特定主题或领域进行高效、深度数据爬取。如抖音、社保掌上通等各类 APP 可以将聚焦网络爬虫的爬取主题设置为“公民个人信息”来收集全站所有或特定的公民个人信息。三是增量式网络爬虫。该类爬虫主要用于爬取网页更新内容，在一定程度上确保爬取网页是尽可能新的网页，具有时效性。主要应用于商业搜索引擎。四是深层网络爬虫。它能够深入网页爬取隐藏数据，适用于数据深度分析。比如可在房地产网站中爬取楼盘经纬度等具体地理数据。

为确保网站的正常运行和数据竞争优势，站点往往会采取两种措施来强化其数据安全保障。其一，采取“协议共识”机制，该机制主要方式为在网站根目录下放置特定文件，文件中详细规定了哪些数据不可抓取。当网络爬虫访问网站时，它首先会检查该页面是否包含这样的协议文件。若存在，爬虫就必

须在规定的范围内进行数据访问；若不存在，爬虫则可无限制地访问页面数据。但是，这种协议共识机制并不具备法律强制效力，它主要依赖于用户自觉遵守。其二，为了进一步提升数据安全保护体系，网站会建立反爬虫技术壁垒，包括身份验证、IP 地址封锁等多种技术手段，旨在有效阻止任何未经授权的数据访问行为。

2.2. 网络爬虫刑法规制必要性

数据爬取行为代表着搜索模式的高效革新，其带来便利的同时，也会带来风险。当风险转化为现实危害，就需要刑法介入。在互联网上，信息和内容主要以数据作为存在和传播的媒介。数据爬取行为一方面可能因为未经授权而侵害数据所承载的相应权益，另一方面可能导致网站拥堵或过多消耗网站资源而影响计算机系统正常运行。

2.2.1. 扰乱计算机信息系统正常运行

行为人利用爬虫工具可以完成对目标网络的定向、高速搜罗，但部分收集活动可能会突破服务器的安全保护机制，进而影响服务器的运行安全，甚至可能危及信息存储的安全性。数据爬取行为可能会扰乱计算机信息系统功能，其原因在于，网络爬虫作为一种自动化的数据抓取工具，运作效率极高，有可能在短时间内发出大量链接请求，这些请求会集中冲击计算机系统，导致系统无法及时、有效地处理其他正常服务请求。这种行为不仅可能引发服务器的拥堵和服务中断，甚至会导致服务器资源被过度消耗而枯竭。这些都严重干扰了计算机信息系统的正常秩序，并对其整体安全构成破坏。另外，数据爬取行为还存在危害计算机信息系统安全的情况，即作为数据爬取行为组成部分的侵入计算机行为，若未经授权或属于其他非法侵入的情形，则存在危害信息系统安全的风险。因此，在互联网上爬取数据，其恶意行为势必会侵害计算机信息系统安全。

2.2.2. 侵害重要数据承载的相应权益

从上述分析可知，网络爬虫是模拟人类获取网页信息，也就是说网络爬虫只能爬取人类可以访问的页面上的数据。行为人利用爬虫工具获取数据时，虽然更多获得的只是普通的、效能相对低下的数据或信息，但也会收集到诸如公民个人数据信息、国家秘密、商业秘密、作品或录音录像制品等敏感内容。以上的价值性数据一旦被恶意收集乃至利用，可能会对客体造成对应侵权的后果。第一，数据爬取行为可能侵害公民的个人信息自决权。《刑法》第 253 条规定，窃取或者以其他方法非法获取公民个人信息的，构成侵犯公民个人信息罪。那么，未经授权爬取公民个人信息的行为，则可能侵害公民的个人信息自决权，需要负担民事责任甚至刑事责任。第二，数据爬取行为可能侵害著作权。由于具有著作权的作品通常以公开形式在网上呈现，因此较容易成为数据爬取行为的对象。根据《著作权法》第 10 条规定，作品的复制权，即以复印、录音、录像、数字化等方式将作品制作多份的权利，属于著作权中的财产权。利用网络爬虫技术抓取具有著作权作品的行为，本质上属于以数字化方式将数据作品制作多份的复制行为，该行为若未经著作权人允许，则侵害了其著作权。第三，数据爬取行为还可能因不当抓取其他数据而侵害数据安全。如非法获取计算机信息系统数据案件中，利用网络爬虫技术爬取百度网盘分享链接地址和提取码、淘宝或微博等账号密码、cookie 等其他数据的行为，则侵害了数据安全。

3. 数据爬取行为规制困境

现有的刑法规制较为笼统且涵摄范围较分散，在具体的法律适用中法官拥有较大的自由裁量空间，导致对不当数据爬取行为入罪门槛低，“口袋化”罪名滥用的倾向。对数据爬取行为的不当规制也限制了数据作为互联网时代的新一类生产要素的使用。

3.1. 以数据载体安全为法益的定位失衡

从目前司法实践来看，数据爬取行为侵害法益的定位为数据载体安全，导致同一案件不同法律承担责任与否产生矛盾，法秩序不统一以及非法获取计算机信息系统数据罪的“口袋化”。

1、法秩序不统一。根据全国首例涉及“网络爬虫”的刑事判决可看出，当前司法实践中对数据和信息进行了明确区分，认为数据保密性指作为信息载体的数据的保密性，而非信息内容本身的保密性。因此即使本案中视频的信息内容处于公开状态，也不代表视频数据处于公开状态，不能认为被告单位没有侵害数据保密性，据此，本案被告单位采用网络爬虫技术，爬取信息公开但代码保密的视频数据的行为具有违法性[6]。然而，值得注意的是，《民法典》《著作权法》以及《个人信息保护法》均明文规定，未经授权而爬取公开信息的行为不承担民事责任或行政责任。这就导致将不负民事责任或行政责任的数据爬取行为认定为犯罪，显然与前置法相矛盾，违反了法秩序统一的原理。

2、非法获取计算机信息系统数据罪“口袋化”。基于以上对数据爬取行为的危害分析可知，此类行为主要侵犯的是数据承载内容的相应权益。然而，若数据载体安全作为保护对象，那么判断重点会浮于行为对象是否为可还原为计算机信息系统数据，从而忽视了对数据所承载的具体法益的深入辨别。有学者对非法获取计算机信息系统数据案件进行了实证分析，发现本罪的犯罪对象几乎涵盖了所有可以在计算机信息系统中存贮的对象，包括，身份认证数据、公民个人数据、网络虚拟数据、数字货币、财产性利益、作品和其他数据等 8 类[7]。以网络虚拟财产、数字货币等数据为例，在当今社会下，公众也默认网络虚拟财产属于个人财产一部分，将虚拟财产解释为刑法范畴内的财物，这一做法并未侵犯国民的预测可能性。因此，对于非法获取此类数据的行为应当根据其具体性质构成相应的财产犯罪，而非一概而论地以非法获取计算机信息系统数据罪进行定罪。综上所述，将不当获取具有不同法律属性信息的行为统一认定为非法获取计算机信息系统数据罪，将进一步加剧该罪名的“口袋化”倾向，不利于法律的精确适用与公正裁判。

3.2. 以刑法作为主要规制途径的理念失常

从当前司法实践来看，对数据爬取行为的规制存在过度依赖刑法来保障数据安全的理念倾向。这一做法降低了数据爬取行为的入罪门槛并且混淆了数据爬取行为与数据利用行为。

1、情节显著轻微的行为入罪。当前采取刑罚手段惩治非法获取数据行为，明显存在将情节显著轻微的非法数据行为一律入罪的问题。从规制数据爬取行为的司法实践来看，“我国对网络爬虫的司法实践规制在强化数据安全理念下日趋‘严厉’，逐步由民事领域转向刑事领域。”[3]以全国首例爬虫入刑案为例，¹由于“爬虫”抓取的视频数据不具有著作权，因此不能被认定为侵犯著作权罪，但被认定为非法获取计算机信息系统数据罪。但是本案的爬取对象属于公开信息范畴，且造成的实际损失较小，被害单位仅损失技术服务费人民币 2 万元，这一数额远低于乐视网诉电视猫非法盗链侵权及不正当竞争纠纷案中法院认定的乐视网损失 50 万元。而后一案件中爬取行为仅被认定为网络信息传播权，负民事责任。²由此可见，若以“刑法先行”的方式规制数据爬取行为容易导致情节显著轻微危害不大的行为被不当入罪，这显然违反了《刑法》第 13 条的规定。

2、混淆数据爬取行为与数据利用行为。数据爬取行为涉及的犯罪存在两种行为构造，即以数据爬取行为为实质的犯罪施行行为构造和以数据爬取行为为手段行为的构造。过度依赖刑罚手段规制数据爬取行为，导致对数据犯罪行为的刑法规制前置化，即在数据被不当使用之前，则基于数据可能被不当使用的风险而规制数据获取行为，使数据利用行为的独立性相对被忽视，导致将利用数据爬取行为实施的其

¹案号：(2017)京 0108 刑初 2384 号载中国裁判文书网，访问时间：2024 年 8 月 20 日。

²案号：(2020)京民再 142 号载中国裁判文书网，访问时间：2024 年 8 月 20 日。

他犯罪的行为认定为非法获取计算机信息系统数据罪。但大多数案件中,非法获取数据行为与数据利用行为存在牵连关系,由于很难被害人的损失的直接行为为数据爬取行为还是非法利用行为。因此,法院会基于非法爬取行为认定被告成立非法获取计算机信息系统数据罪,以达到规制不当利用数据的目的。

3.3. 以平台授权取代法律授权的逻辑失当

此做法实际上是以数据平台的授权与否以及授权大小对数据爬取行为定罪量刑。以平台授权等同于法律授权,不仅导致入罪门槛低,忽视平台未尽到安全保护义务的过错,加重了行为人应承担的责任,还会抑制数据产业发展。

1、降低了入罪门槛,不当加重行为人责任承担。规制数据爬取行为的罪名均为法定犯,行为具有前置法违法性为认定犯罪的首要要求,而网络爬虫的违法性判断取决于是否“未经授权”。但是刑法中是否“违反国家规定”主要依据《网络安全法》第27条,即“任何个人和组织不得从事非法入侵他人网络、干扰他人网络正常功能、窃取网络数据等危害网络安全的活动。”其中“窃取网络数据”是指未经数据控制者允许,获取网络中存储、传输、处理数据的行为。因此,实务界均认可通过判断行为人依据控制者的授权和授权范围来认定行为违法性。但是这也意味着数据网站可根据自身利益来设定网络爬虫的权利义务。可能将在前置法上不具有违法性但违背数据控制者意愿的行为认定为犯罪,使数据爬取行为普遍入罪。除此之外,数据控制者本身具有数据安全保护义务,应当采取必要技术措施保障数据安全。而依据数据控制者授权来判断爬虫行为违法性的做法,相当于让数据爬取者承担所有数据安全风险,而忽略了数据控制者在数据安全事故中应负的责任,不当加重了数据爬取者的责任。

2、抑制数据产业发展。当前用“善意”“恶意”来区分数据爬取行为,而“善意”与“恶意”的区别则取决于是否遵循爬虫协议,无视或违反爬虫协议的行为称为“恶意爬虫”,并将其规定为犯罪。这就意味着用刑罚的手段巩固了数据统治者对数据的垄断,抑制了数据的流动。理论上,数据虽散布于整个互联网空间,实际上却被政府及大型互联网企业控制。尤其是,作为信息技术的提供商,互联网企业推动了数字控制型社会的形成,其凭借技术优势在各自领域形成了事实上的数据垄断,并通过技术迭代升级不断巩固自身的“数据寡头”地位^[8]。大数据时代,数据作为获取知识的原材料,具有数据寡头地位的数据平台挖掘数据价值具有正当性,而普通公众却无法获取公开信息的数据原料,这也就阻断了普通公众从原始数据中挖掘不同层面信息进行创新的可能性,如此恶性循环会加剧数字鸿沟^[1]。

4. 数据爬取行为刑法规制路径探析

4.1. 以数据内容安全为法益,加强现行法律适用

通过上述分析可知,数据爬取行为法益应为数据承载的实质内容价值而非作为信息载体的数据的安全。在此基础上应加强现行刑法罪名的适用。其一,侵犯计算机信息系统安全罪。《刑法》第285条属于选择性罪名,针对不同目的、不同形态的实行行为,可分别产生不同的犯罪构成。从罪状来看,恶意爬取数据的目的仍是获得数据而非单纯对“安全认证系统”或者对授权范围的突破,具体而言,此类行为表现为行为人利用“爬虫”程序非法侵入计算机信息系统后获取数据,或采用其他技术手段非法获取数据。其主要涉及的罪名大多是“非法获取计算机信息系统数据罪”。其二,侵犯公民个人信息罪。不论价值与否、是否关涉隐私,互联网数据在客观上纠合了大量可识别的个人信息³,侵犯公民个人信息罪的

³根据《最高人民法院 最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》:“以电子或者其他方式记录的能够单独或者与其他信息结合识别特定自然人身份或者反映特定自然人活动情况的各种信息,包括姓名、身份证件号码、通信通讯联系方式、住址、账号密码、财产状况、行踪轨迹等。”;根据《中华人民共和国民法典》第一千零三十四条第二款:“个人信息是以电子或者其他方式记录的能够单独或者与其他信息结合识别特定自然人的各种信息,包括自然人的姓名、出生日期、身份证件号码、生物识别信息、住址、电话号码、电子邮箱、健康信息、行踪信息等。”

犯罪客观方面主要包括两种行为类型：一是向他人出售或者提供公民个人信息的行为，二是窃取或以其他方式非法获取公民个人信息的行为。有学者认为，“侵犯公民个人信息罪的前置法益内容是个人信息权益，而刑法法益则是其他人身与财产安全。”[9]恶意数据爬取行为止于获取信息的一步，符合第二种行为范式，基本的犯罪构成为行为人违反了国家有关许可规定，非法爬取、收集公民个人信息，行为属于窃取或以其他方式非法获取，达到了情节严重的程度。其三，侵犯著作权罪的规制路径。根据《中华人民共和国著作权法》，构成作品需要满足“可复制性”和“独创性”双重要件，其中“独创性”是判定网络内容是否构成作品的核心认定要件，也是实践中频频出现的争议焦点。恶意爬取数据构成侵犯著作权罪行为经常体现为行为人利用爬虫工具非法接入搜索引擎、视频聚合平台等网络产品，具体以“网页转码”和“加框链接”为典型[10]。入罪的另一项重要前提是“未经著作权人许可”，而在广泛爬取网页和后续使用作品时，一般难以事先获得著作权人的许可或出现权利人分散的客观现象，两项前提共同完成对恶意爬取行为的侵犯著作权罪入罪评价。此外，侵犯著作权罪的犯罪形态为结果犯，恶意爬取他人著作并不能单独构成此罪，还需要行为人“以营利为目的”，并且存在获取数据后的数据使用行为，如通过复制发行、通过信息网络向公众传播著作权人作品、录音录像、表演或故意避开、破坏保护著作权及有关权利的技术措施等方式，导致“违法所得数额较大或者有其他严重情节”[11]。

4.2. 整体把握数据爬取行为违法性，平衡数据保护与数据垄断

将网络爬虫的违法性判断等同于数据控制者的授权，实质上是将网络爬虫的违法性判断委托给数据控制者，不仅在违法性判断上有失公允，还加剧了数据垄断的局面。整体法秩序上判断数据爬取行为的违法性，找到数据保护与反垄断之间的平衡点，是爬取数据行为刑法规制的一个重点。首先，以前置法违法为前提，这是罪刑法定原则的要求。从现行刑法规制数据爬取行为的罪名来看，都是法定犯，也就是为了国家行政取缔目的特别设立的犯罪。法定犯具有行政违法与刑事违法双重违法性，因此，其行政违法的确认是判断法定犯是否构成犯罪的前提条件。数据爬取行为相关罪名中，通常涉及军事机密、国家秘密、公民个人信息等规范性构成要件要素，而这些要素必须根据相关法律、法规作出评价。其次，数据持有者监管过失责任。《网络安全法》第21条规定了数据控制者应采取技术措施和其他必要措施保护网络安全。若数据控制者对于危害数据安全结果的发生未尽到数据安全保护义务的责任，数据爬取行为人的期待可能性降低，应当对其从宽处理。最后，许多商业模式在正式运行时，往往会将多种数据综合运用。然后，这些数据是否能及时获取，还需通过具体商业模式分析确定。尤其是稀缺性数据，因此不可替代性，立法者应出于反垄断的考量，引入“必要设施理论”，要求稀缺数据持有者承担开放数据的义务。这也是推动互联网发展的一项有效法律举措。

4.3. 以刑法为保障法的保护体系，合理划定法律红线

刑法的过度适用导致将情节显著轻微的数据爬取行为一律入罪，针对此问题，应妥善适用《刑法》第13条“但书”情节。对于未侵害国家秘密和情报、信用卡信息、商业秘密、公民个人信息，仅造成竞争利益损失的情形按照不正当竞争案件处理，对数据爬取行为确有刑法规制必要时方作为犯罪处理。以此构建完善的民事责任、行政责任以及刑事责任体系，保持刑事打击与民事救济之间的平衡，保护互联网平台的合法权益、维护网络空间的正常秩序。其一，非法获取计算机信息系统罪在非法获取数据罪体系中属于兜底罪名，非法爬取刑法重点保护的国家秘密和情报、著作权、公民个人信息、信用卡信息、商业秘密、军事秘密以外的其他数据，侵害信息保密性时，可以认定为本罪。而本罪的低入罪门槛也给本罪带来了高适用率，使本罪沦为“口袋罪”。基于刑法谦抑性精神，按照不正当竞争纠纷案件处理，足以解决“爬虫”行为人和被爬取数据平台之间的矛盾[12]。其二，刑法在网络安全领域介入时间的提前，

使得数据爬取行为与数据利用行为相混淆,利用爬取的数据实施其他犯罪的行为也被认定为非法获取数据类犯罪。对于数据利用行为不构成犯罪或难以构成犯罪的情况,基于刑法谦抑性精神,只有在整体犯罪行为社会危害性较大时应受刑法处罚的,可以考虑适用非法获取计算机信息系统数据罪处罚非法数据爬取行为。

参考文献

- [1] 李标,孙琨,孙根紧.数据要素参与收入分配:理论分析、事实依据与实践路径[J].改革,2022(3):66-76.
- [2] 郭春镇,勇琪.算法的程序正义[J].中国政法大学学报,2023,93(1):164-180.
- [3] 杨志琼.数据时代网络爬虫的刑法规制[J].比较法研究,2020(4):185-200.
- [4] 周德懋,李舟军.高性能网络爬虫:研究综述[J].计算机科学,2009,36(8):26-29+53.
- [5] 韦玮.精通Python网络爬虫:核心技术、框架与项目实战[M].北京:机械工业出版社,2017.
- [6] 游涛,计莉卉.使用网络爬虫获取数据行为的刑事责任认定——以“晟品公司”非法获取计算机信息系统数据罪为视角[J].法律适用,2019(10):3-10.
- [7] 杨志琼.非法获取计算机信息系统数据罪“口袋化”的实证分析及其处理路径[J].法学评论,2018,36(6):163-174.
- [8] 单勇.跨越“数字鸿沟”:技术治理的非均衡性社会参与应对[J].中国特色社会主义研究,2019(5):68-75+82.
- [9] 刘浩.侵犯公民个人信息罪的法益构造及其规范解释[J].环球法律评论,2023,45(3):163-179.
- [10] 麦买提·乌斯曼,杨立敏.“算法推送”与网络服务提供者著作权侵权刑事责任规范性重构——从“实际作用”转向“规范能力”[J].重庆理工大学学报(社会科学),2021,35(8):147-159.
- [11] 詹红星,王李.网络爬虫行为的入罪路径及其限度[J].北京邮电大学学报(社会科学版),2023,25(1):94-103.
- [12] 蒋巍.恶意数据爬取行为的刑法规制研究[J].学术论坛,2020,43(3):48-54.