

网络爬虫行为刑法规制问题研究

张 涛

贵州大学法学院, 贵州 贵阳

收稿日期: 2023年3月6日; 录用日期: 2023年3月17日; 发布日期: 2023年5月19日

摘 要

网络爬虫行为, 是行为主体使用网络爬虫技术进行数据搜集、整理的行为。网络爬虫行为存在善恶之分。网络爬虫行为是否入刑, 关键在于数据所有者、控制者是否授权, 若获得授权则无需遭受法律非难, 反之则应纳入法律评价范围, 具体从行为不法和对象不法两方面进行展开。行为不法层面, 违背Robots协议承担相应民事责任、恶意爬虫行为受到刑法约束; 对象不法层面, 根据数据性质决定行为主体是否承担责任, 爬取公开的普通数据无需承担任何责任、爬取保护的数据需承担相应民事、刑事责任。网络爬虫行为关涉罪名主要为: 侵犯公民个人信息罪、侵犯著作权罪、非法获取计算机信息系统数据罪以及破坏计算机信息系统罪。

关键词

网络爬虫行为, 主观违法性, 授权, 行为不法, 对象不法

Research on the Criminal Regulation of Internet Crawler Behavior

Tao Zhang

School of Law, Guizhou University, Guiyang Guizhou

Received: Mar. 6th, 2023; accepted: Mar. 17th, 2023; published: May 19th, 2023

Abstract

Web crawler behavior is the behavior of the subject using web crawler technology for data collection and collation. There are good and evil in web crawler behavior. Whether the behavior of network crawler is criminalized depends on whether the data owner and controller are authorized. If authorized, it does not need to suffer legal reproach. Otherwise, it should be included in the scope of legal evaluation, which is carried out from two aspects: illegal behavior and illegal object. At the level of illegal behavior, violating the robots agreement to bear the corresponding civil liability,

文章引用: 张涛. 网络爬虫行为刑法规制问题研究[J]. 法学, 2023, 11(3): 1054-1060.

DOI: 10.12677/ojls.2023.113150

malicious crawler behavior is subject to criminal law; at the illegal level of the object, according to the nature of the data, it is determined whether the subject of the behavior is responsible. There is no need to bear any responsibility for crawling the public ordinary data, and the data to be protected should bear the corresponding civil and criminal responsibility. The crimes involved in the network crawler behavior are mainly: the crime of infringing citizens' personal information, the crime of infringing copyright, the crime of illegally obtaining computer information system data and the crime of destroying computer information system.

Keywords

Internet Crawler Behavior, Subjective Illegality, Authorization, Unlawful Conduct, Illegal Object

Copyright © 2023 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. 问题的提出

大数据背景下,网络爬虫逐渐成为热潮。网络爬虫作为一种技术无善恶之念,但在居心叵测之人操纵下,而演变为不法分子谋取暴利、实施犯罪的工具。作为全国首例以爬虫技术侵入计算机系统的晟品网络科技有限公司(以下简称“晟品公司”)案中,晟品公司员工张某、宋某等人利用技术手段抓取并破解字节跳动网络技术有限公司(以下简称“字节跳动公司”)服务器中储存的数据,造成字节跳动公司遭受额外经济损失,最终法院判决晟品公司及实施不法行为的相关员工构成非法获取计算机信息系统罪。¹此案在理论和实务界一石激起千层浪,由此引发众多学者对网络爬虫带来的法律风险进行讨论。

有鉴于此,笔者在中国裁判文书网中,以“网络爬虫”、“网络爬虫行为”为关键词检索发现,截止目前裁判文书网中总共存在79篇相关主题的法律文书。其中78篇判决书,1篇裁定书,民事案件存在57件,刑事案件10件,行政案件7件。刑事案件涉及的罪名主要为侵犯公民个人信息罪、侵犯著作权罪、非法获取计算机信息系统数据罪以及破坏计算机信息系统罪,民事案件涉及的主要为侵犯著作权、不正当竞争等行为。由此可以看出,网络爬虫带来的法律风险集中存在于个人信息类犯罪、计算机数据系统类犯罪中且以民事案件为主,但网络爬虫带来的刑事风险也不可忽视。笔者认为有必要对网络爬虫行为定性,为网络爬虫行为的刑法规制提供思路。

2. 网络爬虫行为的认定:有善恶之分

网络爬虫,也称网络机器人、网络蜘蛛,是通过技术手段代替人类在网络空间中搜寻、获取以及整理信息的程序或者脚本。网络爬虫行为,是行为主体通过客户端、终端等载体使用网络爬虫技术进行数据搜集、整理的行为。由此可见,根据使用者的主观意思判断网络爬虫行为有善恶之分。比如,晟品公司利用自身掌握的网络爬虫爬取字节跳动公司的存储数据,该行为导致字节跳动公司遭受较大经济损失,这便是恶意网络爬虫行为;利用网络爬虫,收集、整合互联网中公开的教学资源,供行为人学习使用,既不侵害数据所有者的利益,也有利行为人自身素质提升,这便是善意网络爬虫行为。准确判断涉网络爬虫案件行为主体的主观是善意还是恶意应从以下方面进行考察:

其一,行为主体对网络爬虫技术的熟练程度。若行为人从事互联网行业并且对爬虫技术熟练掌握,

¹ 参见北京市海淀区人民法院(2017)京0108刑初2384号刑事判决书。

实施网络爬虫行为，在主观上便存在故意。比如，晟品公司作为一家科技公司，其员工张某、宋某等人有着丰富的计算机学方面知识且长期从事互联网行业，对爬虫技术的风险有着清晰认知，却仍然伙同他人非法爬取字节跳动公司的数据，由此判断张某、宋某等人主观上是恶意。

其二，行为主体是否设置特殊的爬取模式。若行为人开发、使用的爬虫技术能够自动避开数据管理者设置的反爬措施(包括屏蔽 IP、躲避身份验证等)以爬取数据信息，那么应该推定其为恶意。

其三，行为主体是否爬取特定类型数据。特定类型数据主要针对公民个人基本信息，比如电话号码、家庭住址、照片等或者是著作权或者是商业秘密，对此类特定数据的分析也能推断出行为主体主观上为恶意。比如，被告人周华通过个人编写的网络爬虫侵入 XX 学院教务系统，非法爬取学生照片、姓名以及身份证号等数据信息，并将存储该信息的服务器网址公布于众，以此来检验自己开发的爬虫程序能否有效运行，最终法院判决其构成侵犯公民个人信息罪。²

其四，行为主体是否采取防范风险举措。行为主体对网络爬虫赋有防范风险义务，对网络爬虫的爬取对象、爬取范围应该明确限定，同时应该履行风险防控的作为义务[1]并实时监控网络爬虫的状态，避免网络爬虫对其他权利人的合法权益造成不应有的损害。

行为主体实行良善的网络爬虫行为既不侵害其他权利人合法权益，亦能够高效整理网络数据信息，能最大限度发挥网络爬虫的便捷优势，实现互利共赢。但在复杂的市场环境下，部分行为主体经不住暴利的诱惑或者是法治意识淡薄，便会实施恶意网络爬虫行为。比如，晟品公司利用爬虫技术非法获取数据以获得高额收益最终获罪判刑，此种便属于经不住利益诱惑之例；被告人周华为测试个人开发的爬虫程序，而侵入某学院教务系统爬取大量学生的个人信息，最终难逃法网获罪判刑，但法官因其法治意识淡薄而从轻处理。

3. 网络爬虫行为刑法规制的困境及必要性

3.1. 网络爬虫行为刑法规制面临的困境

3.1.1. 网络爬虫行为入罪标准之争

网络爬虫行为存在善恶之分，恶意的网络爬虫行为应当受到法律规制，善意的网络爬虫行为不在 Robots 协议和法律评价的范围内。有的研究者认为，“当前的刑事规制过于注重网络爬虫的违法性侧面，不利于网络爬虫技术的应用与发展[2]。”应该将关注的重点放在网络爬虫行为的合法性层面。有的研究者认为，网络爬虫行为入罪应该通过形式判断实现形式入罪与实质出罪双重机制[3]，即使网络爬虫行为具备非法性，也实施了非法爬取数据的行为，但是该数据是无法识别、无法使用的“没有任何价值”的，在形式上虽具备入罪条件，但实质上并未侵害法益，故而应该实质出罪。同时，也有研究者指出，应该从行为不法、对象不法层面来判断是否将爬虫行为入罪。网络爬虫行为入罪标准不一，往往会导致实务中司法人员定罪量刑不规范，难以切实贯彻“罪法定原则”，不利司法公正。

3.1.2. 立法滞后

总体而言，我国涉网络爬虫相关问题立法滞后[4]。近年来，个人信息保护法、数据安全法、网络安全法、密码法以及网络安全审查办法等法律法规的颁布为网络空间治理提供了多维度保护机制，营造了良好的网络运营环境。但同时网络空间发展日新月异，不断涌现出侵害数据安全新型法益行为[5]，面对此种现状，现行笼统、抽象的法律法规往往难以有效规制，特别是利用网络爬虫非法侵入计算机系统爬取数据与提供侵入、非法控制计算机信息系统的程序、工具罪二者的认定便存在争议[6]。只有在立法或司法解释中明确规定相关涉网络爬虫行为的罪名，才能有效规制现实不法状况。

² 参见湖南省 XX 市鹤城区人民法院(2019)湘 1202 刑初 530 号刑事判决书。

3.1.3. 司法实务面临困境

上文谈及，网络爬虫行为存在善恶之分，在定罪量刑时对行为主体主观意图的判断便至关重要，主观违法性便需要结合上文所言的四条标准来判断，与此同时，司法机关在实务中对涉网络爬虫案件也面临难取证、难归责定罪问题。比如，林镇平等人通过网络爬虫技术非法获取北京某公司计算机信息系统数据造成该公司遭受巨额经济损失，最终林镇平等人被法院判处非法获取计算机信息系统数据罪。³该案中，由于林镇平等人采用破解验证码、绕开挑战登录等方式逃避反爬措施并且隐藏爬虫痕迹，对司法机关而言，搜集证据便更具挑战，与此同时，该案涉及到单位犯罪、共同犯罪以及自首等情节，对办案人员准确定罪量刑也提出了更高要求。

3.2. 网络爬虫行为刑法规制必要性分析

刑法作为各部门法的“保护法”，最具严厉性。著作权法、个人信息保护法以及网络安全法等法律法规在一定程度上能保障权利人的合法权利，但当行为人合法权利遭受严重侵害时，便需刑法来予以救济以恢复权利圆满状态，与此同时，大量无视规则、任意爬取的行为不断发生^[7]，故而需要利用刑法手段对网络爬虫行为进行规制。

3.2.1. 有利于保护数据所有者、控制者的合法权益

恶意行为主体利用网络爬虫非法爬取网站、计算机系统中储存的数据信息以实现谋取经济利益或测试爬虫程序是否有序运营等目的，侵害了原数据所有者、控制者的合法权益。有研究数据表明，网络中存在三分之二的信息获取是恶意的⁴，即是说，网站中多数信息是从非正规渠道获得，这样将会导致恶性竞争等行为的产生。比如，杭州阿里巴巴广告有限公司、阿里巴巴(中国)网络技术有限公司等与南京码注网络科技有限公司(简称“南京公司”)等不正当竞争纠纷案中，被告南京公司利用网络爬虫技术抓取原告(杭州阿里巴巴公司、中国阿里巴巴公司)在网站中公开显示的用户“联系地址”、“联系电话”、“联系人”以及“诚信通使用年限”等信息并直接在其运营的网站上显示，其行为严重损害了原告合法权益并受到相应的处罚。⁵针对此种恶意爬取行为造成严重后果，应将其入罪并充分发挥刑法的威慑功能，以便更好保护数据归属。

3.2.2. 有利于形成良好的网络空间生态环境

生态环境已不再限定为物理世界的生态系统。随着互联网发展，人们广泛参与网络互动，逐渐形成了网络空间生态。良好的网络空间生态环境是网络世界有序发展的前提。不论是南京公司非法爬取阿里巴巴公司的用户数据亦或是晟品公司非法爬取并破解字节跳动公司的存储数据，虽然前者为民事案件，后者为刑事案件，二者在案件性质上有所不同，但是二者客观上对互联网行业正常的交流活动造成了不同程度的影响，对有序的网络空间生态造成了破坏。对此，笔者认为应加大对涉案人员的处罚力度以充分发挥刑法的惩罚功能，营造有序的网络生态环境。

3.2.3. 有利于技术创新发展

科学技术是第一生产力。网络爬虫作为一项技术，在现代化科技进程中具有重要的地位。通过强有力的法律来规制非法使用网络爬虫行为，一方面将严厉打击不法使用者，使其遭受所犯不法行为的“恶”；从另一方面而言，国家使用法律手段维护了合法使用网络爬虫行为者以及数据信息所有者、控制者的权益，鼓励并推动网络数据资源的开放共享，大力支持利用网络爬虫技术进行合法数据采集、整理，通过这一系列有力举措营造出良好的网络环境，让网络爬虫开发者更有信心、更加坚定网络爬虫技术未来具

³ 参见北京市朝阳区人民法院(2020)京 0105 刑初 2594 号刑事判决书。

⁴ 参见 http://www.cac.gov.cn/2019-06/16/c_1124630015.htm?from=singlemessage，最后访问日期为 2023 年 2 月 11 日。

⁵ 参见浙江省杭州市滨江区人民法院(2019)浙 0108 民初 5049 号民事判决书。

备无限发展的潜力。安定有序的互联网空间生态，是人们进行技术创造的前提，技术创新发展对推动互联网行业大有裨益。

4. 网络爬虫行为刑法规制的路径

随着社会将数字技术应用于更多有益目的，必须在法律层面明确区分网络爬虫的合法使用和违法犯罪[8]。网络爬虫行为是否入罪，核心是行为主体的行为是否获得数据所有者、控制者赋权，具体表现在行为层面针对恶意的网络爬虫行为，在对象层面根据爬取数据的性质来判断是否入刑。

4.1. 行为不法层面

4.1.1. 违反 Robots 协议承担相应民事责任

Robots 协议的英文全称为 Robots Exclusion Protocol，翻译为机器人排除协议，或者称为爬虫协议、机器人协议，是指网站所有者通过一个置于网站根目录下的文本文件，即 robots.txt，告知搜索引擎的网络机器人(或称网络爬虫)哪些网页不应被抓取，哪些网页可以抓取，其本质上是受访网站与搜索引擎之间的一种交互方式。⁶ Robots 协议是一种网络行业惯例，是一种技术规范，一方面要求执行搜索任务的网络爬虫遵守受访网站的协议，同时也要求受访网站所设置的协议本身是合理的，不违背“信息共享”的初衷。因此，行为主体实施的爬虫行为违背了该协议应该承担侵害著作权责任、反不正当竞争等责任。例如，百度公司 VS. 奇虎 360 案，百度公司违背 Robots 协议限制奇虎公司正常抓取其网页信息，导致奇虎公司部分搜索引擎无法使用，损害了奇虎公司的合法权益，构成不正当竞争，需承担相应不正当竞争法责任。⁷

4.1.2. 恶意的网络爬虫行为承担相应刑事责任

行为主体实施恶意的网络爬虫行为主要表现在是否故意避开对方反爬措施、是否使用技术手段突破对方设置的反爬措施。我国涉数据犯罪的相关立法和司法解释认为行为主体实施网络爬虫行为时，只有突破、避让、绕过[9]网站设置的技术障碍时才能纳入刑法评价的范围。比如，我国《刑法》规定的非法获取计算机信息系统数据罪的构成要件为“采用侵入或其他技术手段”，以及《关于办理危害计算机信息系统安全刑事案件应用法律若干问题的解释》中相关条款也强调网络爬虫行为必须具备能够突破、避开计算机技术屏障的功能。司法实践中，行为主体实行恶意的网络爬虫行为对其他网站、计算机系统数据进行数据爬取，以达到自己不法目的。比如，被告单位厦门房麦网络科技有限公司通过网络爬虫程序，采用破解验证码、隐匿爬虫痕迹等方式非法获取北京某技术公司(原告)经营的某网站房地产数据资料并解密、整合后供本公司使用以抢占市场先机，致使原告遭受严重经济损失，最终法院判决被告因恶意使用网络爬虫技术严重损害他人利益，构成非法获取计算机信息系统数据罪。⁸

4.2. 对象不法层面

网络爬虫爬取的对象是网络数据。根据我国刑法相关规定，刑法评价的“数据”仅仅是计算机信息系统中能够受到访问控制的数据，该种数据依附于计算机信息系统，存在该系统内部[10]。但是随着网络技术的发展，司法实务中出现的“数据”涵盖范围已超越刑法评价的“数据”范围，包括了计算机信息系统外部数据，比如网页的浏览记录、搜索痕迹等。数据本身具备价值性、公共性以及多样性特征，包括公开的和非公开的数据、可访问的和不可访问的数据等类型，故而针对数据应该区别保护。

⁶ 参见北京市高级人民法院(2017)京民终 487 号民事判决书。

⁷ 同注释 6。

⁸ 同注释 3。

4.2.1. 爬取开放的普通数据免责

开放数据是指任何人可在任何时间、地点访问或使用的数据^[11]。意味着任何人均有权利利用开放数据，亦即权利人放弃了对数据的掌控权。行为人爬取开放的数据根据民事法律中的被害人同意理论可免责。爬取开放的数据在刑法领域也可根据“权利人同意”理论免责。就开放的个人数据而言，权利人需对开放的个人数据持容忍义务，不得因他人的合法收集、利用而主张侵权并要求赔偿。就开放的公司数据而言，权利人应该允许他人合理收集、使用，而不得限制或禁止其合法利用，否则将可能承担相关责任。比如，百度公司 VS. 奇虎 360 案中，百度公司禁止奇虎公司爬取其公开的网页信息，致使奇虎公司遭受损失，法院判决百度公司败诉并承担相应的赔偿责任。

4.2.2. 爬取保护的数据将担责

保护的数据，是指任何人未经权利人或管理人许可而以任何途径、方式收集或整理的信息。行为人若需使用保护的数据，需经权利人的授权，未经权利人授权而使用网络爬虫抓取数据将承担相应的责任。

1) 承担民事方面责任

其一，可能承担知识产权侵权责任。笔者在中国裁判文书网中检索发现，涉“网络爬虫”的 57 件民事案件中，著作权侵权案件达 33 件。由此可见，在民事领域著作权侵权已成为网络爬虫行为的主要侵权类型。著作权侵权案件包括直接侵权行为和间接侵权行为。其二，可能承担不正当竞争法责任。反不正当竞争是现实众多企业寻求的一个有利规则模式。通过反不正当竞争规则，相关主体能获得更可观的利益，这也驱使更多企业利用爬虫抓取数据信息以抢占市场先机。这也意味着相关行为主体将承担更多不正当竞争法律责任。

2) 承担刑事方面责任

行为主体实施网络爬取行为情节严重可能触犯以下罪名并承担相应刑事责任：

其一，构成侵犯著作权罪。根据我国刑法相关规定，以营利为目的，有下列侵犯著作权或者与著作权有关的权利的情形之一，违法所得数额巨大或者有其他特别严重情节的，处三年以上十年以下有期徒刑，并处罚金：未经著作权人许可，复制发行、通过信息网络向公众传播其文字作品、音乐、美术、视听作品、计算机软件及法律、行政法规规定的其他作品的；……被告人肖俊以营利为目的，利用爬虫程序自动爬取互联网中各个小说网页链接、小说名称、小说作者、小说简介、小说图片等信息并分类存放在腾讯云服务器中且不间断更新抓取内容并根据读者的点击量获取推广费，构成侵犯著作权罪。⁹

其二，构成侵犯公民个人信息罪。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》相关规定，公民的个人信息，是指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。行为主体通过网络爬虫等技术手段非法爬取公民个人信息，以获取巨额利益，情节严重，应当受到刑法规制。比如，马适之等人利用网络爬虫程序获取互联网中的手机号、网址和登录时间以及“极速数据魔方”软件等数据信息并转卖给他人，谋取了巨额利益，司法机关认为其利用网络爬虫技术非法获取公民个人信息构成侵犯公民个人信息罪。¹⁰

其三，构成非法获取计算机信息系统数据罪。根据刑法相关规定，此罪中的计算机信息系统包括国家事务、国防建设、尖端科学技术领域的计算机信息系统，也包括企业、事业单位等计算机信息系统。行为主体利用爬虫侵入计算机信息系统爬取数据，造成被侵入的计算机信息系统数据流失，不论其是否获利，都应该受到刑法规制。

⁹ 参见江西省永丰县人民法院(2022)赣 0825 刑初 2 号刑事判决书。

¹⁰ 参见湖北省宜昌市中级人民法院(2018)鄂 05 刑终 365 号刑事判决书。

其四，构成破坏计算机信息系统罪。行为主体违反相关规定，利用网络爬虫对计算机信息系统功能进行删除、修改、增加、干扰或者对计算机信息系统中存储、处理或者传输的数据和应用程序进行删除、修改、增加的操作或者故意制作、传播计算机病毒等破坏性程序，影响计算机系统正常运行，将构成破坏计算机信息系统罪。比如，被告人杨杰明、张国栋利用网络爬虫软件爬取居住证系统数据，该爬虫软件自动攻击服务器，造成服务器阻塞，导致为5万以上有效注册用户提供服务的居住证系统1小时以上不能正常运行，后果特别严重。经司法机关调查取证，判定其构成破坏计算机信息系统罪。¹¹

5. 结语

在科学技术飞速发展的今天，“数据是网络的核心，而不是文档或者人，并且这些数据都是可以被机器识别处理的数据，因此，用户从网络上获取信息就像查询数据库一样容易，而不必掌握各网站的数据组织架构”[12]。网络爬虫作为一项抓取数据的技术手段，具备中立性。实施网络爬虫行为的行为主体应意识到网络爬虫的风险并对其采取相应防控措施，同时积极履行作为义务减少、避免因网络爬虫行为给他人造成损失。公民作为网络世界的参与者，要提高个人法治意识，增强法治观念，避免因法治观念淡薄实施恶意网络爬取行为而受到法律制裁，与此同时，面对激烈的市场竞争环境，市场活动参与者应坚守道德底线、法律红线，为营造良好的市场环境尽一份力。理论研究者、司法工作者在立足既有理论、实践情况下，应明确网络爬虫行为入罪标准、出罪机制，实现规范定罪、合理量刑，维护法律的公平正义。

参考文献

- [1] 吴卫. 明确越界网络爬虫行为的刑事处罚边界[N]. 检察日报, 2022-07-20(3).
- [2] 孙禹. 论网络爬虫的刑事合规[J]. 法学杂志, 2022, 43(1): 162-172.
- [3] 刘艳红. 网络爬虫行为的刑事规制研究——以侵犯公民个人信息犯罪为视角[J]. 政治与法律, 2019(11): 16-29.
- [4] 胡立平. 网络爬虫行为的刑法规制[J]. 人民检察, 2022(11): 72-73.
- [5] 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.
- [6] 李淼. 非法提供网络爬虫技术行为的刑法规制[J]. 南京航空航天大学学报(社会科学版), 2022, 24(3): 69-74.
- [7] 旷银. 网络爬虫行为刑事规制问题研究[J]. 贵州警察学院学报, 2022, 34(3): 59-65.
- [8] Din, M.F. (2015) Breaching and Entering: When Data Scraping Should Be a Federal Computer Hacking Crime. *Brooklyn Law Review*, 81, Article No. 11.
- [9] 郭鹏飞. 利用网络爬虫获取个人信息的刑法规制[J]. 人民司法, 2022(10): 21-27.
- [10] 童云峰. 大数据时代网络爬虫行为刑法规制限度研究[J]. 大连理工大学学报(社会科学版), 2022, 43(2): 88-97.
- [11] Borgesius, F.Z., Gray, J. and van Eechoud, M. (2015) Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework. *Berkeley Technology Law Journal*, 30, 2073-2131.
- [12] 刘琼, 任树怀. 论 web3.0 下的信息共享空间[J]. 图书馆, 2011(2): 83-85+96.

¹¹ 参见广东省深圳市南山区人民法院(2019)粤0305刑初193号刑事判决书。