Open Journal of Legal Science 法学, 2025, 13(12), 2698-2706
Published Online December 2025 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2025.1312367

Hans汉斯

网络爬取数据的刑法规制路径展开

王昕宁

长春理工大学法学院，吉林 长春

收稿日期：2025年10月26日；录用日期：2025年11月7日；发布日期：2025年12月2日

摘　要

大数据时代下，网络爬虫技术作为数据获取的核心工具，在推动数字经济发展的同时，其滥用行为也对数据安全、公民权益及市场秩序构成严重威胁。我国刑法通过计算机信息系统犯罪、侵犯公民个人信息犯罪等罪名体系对其进行规制，但司法实践中面临技术中立性与刑事违法性界分模糊、入罪标准不统一、罪名适用混乱等困境。本文结合典型案例与最新立法动态，从网络数据爬取行为的刑事风险谱系出发，剖析当前刑法规制的实践难题，提出完善路径，明确构建“三阶层认定标准”，并在此基础上对“国家有关规定”“侵入”等相关概念进行解释，以期实现数据安全保护与信息产业发展的平衡。

关键词

网络数据爬取，网络爬虫，刑法规制

The Exploration of the Path of Criminal Law Regulation for Online Data Crawling

Xinning Wang

School of Law, Changchun University of Science and Technology, Changchun Jilin

Received: October 26, 2025; accepted: November 7, 2025; published: December 2, 2025

Abstract

In the era of big data, web crawler technology, as a core tool for data acquisition, promotes the development of the digital economy. However, its abuse also poses a serious threat to data security, citizens' rights and interests, and market order. China's Criminal Law regulates such abuse through a system of charges, including crimes against computer information systems and crimes of infringing on citizens' personal information. Nevertheless, judicial practice faces difficulties such as ambiguity in distinguishing between technical neutrality and criminal illegality, inconsistent standards for criminalization, and confusion in the application of charges. Combining typical cases and the

文章引用: 王昕宁. 网络爬取数据的刑法规制路径展开[J]. 法学, 2025, 13(12): 2698-2706.
DOI: 10.12677/ojls.2025.1312367

latest legislative developments, this paper starts from the criminal risk spectrum of online data crawling activities, analyzes the practical difficulties in current criminal law regulation, proposes improvement paths, clearly constructs a "three-tier identification standard", and on this basis interprets relevant concepts such as "relevant national provisions" and "intrusion", with a view to balancing data security protection and the development of the information industry.

Keywords

Online Data Crawling, Web Crawler, Criminal Law Regulation

Copyright © 2025 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

网络爬虫，亦称网络蜘蛛(Web Spider)、网络蚂蚁(Web Ant)、网络机器人(Web Robot)等，是一类依据预设算法规则，自动化完成互联网特定数据浏览、收集与抓取的技术程序。其核心功能在于模拟人类用户的网页访问逻辑及数据提取行为，以此高效获取目标信息，典型具备数据采集高精准性、数据覆盖范围广泛性、技术应用准入门槛较低及对目标网站构成较高防御难度四大技术特征[1]。需明确的是，网络爬虫技术本身具有工具中立性，其功能属性不直接关联法律评价；但网络爬虫行为的实施过程，始终体现行为人的主观意志与目标导向，且可能对数据安全、信息权益及网络秩序产生实质影响，具备法律层面的评价必要性与现实意义。基于此，为防范不当爬虫行为引发的法律风险，维护数字空间的正常秩序与合法权益，有必要依托刑法规范对其予以明确规制。

(一) 爬虫技术的价值二元性

网络爬虫作为自动化数据采集技术工具，通过模拟浏览器 HTTP/HTTPS 请求、解析服务器响应数据，实现海量信息的高效获取，已成为搜索引擎运维、学术研究、市场分析等合法场景的基础性技术支撑。最高人民法院在 2021 年大连某数据平台管理中心与崔某某侵害技术秘密纠纷一案中肯定了爬虫技术的中立性，针对被告提出的爬虫技术本身即为违法技术的抗辩，最高人民法院认为即使爬虫技术曾被用于违法活动，但并不等于该项技术本身具有违法性[1]。需明确的是，技术中立性并不等同于行为正当性。当爬虫行为突破目标系统安全防护机制、过度采集敏感数据时，其性质即从合法工具演变为数据侵权乃至刑事犯罪的载体。全球网络安全企业 Imperva 在《2025 年网络安全报告》中指出，2024 年恶意机器人流量已占全球互联网总流量的 37%，较 2023 年的 32%显著增长；其中，高阶与中阶机器人攻击占所有机器人攻击总量的 55%[2]。攻击者正愈发频繁地运用复杂技术模拟人类访问行为实施恶意活动，导致此类攻击的检测与缓解难度显著提升。

(二) 数字经济下的刑法规制需求

随着《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等前置性法律的颁布施行，我国已初步构建起数据保护领域的系统性法律框架。但实践表明，仅依靠民事制裁与行政监管，对逐步规模化的非法爬取数据行为的规制效能存在明显局限，此类手段不仅难以足够威慑犯罪人，也无法完全覆盖非法爬取行为引发的多重风险。在中国裁判文书网的公开裁判文书数据库中，以“爬虫”为核心检

[1]参见最高人民法院(2021)知民终 1687 号民事判决书。
[2]https://www.imperva.com/resources/lp/2025-Bad-Bot-Report-SCH.pdf, 2025/10/4.

索词进行案例检索，经进一步甄别筛选，排除与“爬虫”相关行为无实质关联的案件后，截至本次检索操作完成之日(2025 年 10 月 4 日)，共筛选出符合检索标准的刑事案件 117 件、民事案件 711 件。该检索数据清晰印证：网络爬虫作为兼具数据获取便捷性与行为风险性的技术工具，其风险防控机制的完善与合法行为边界的界定，亟需依托刑法规范予以明确规制。从司法实践来看，涉爬虫案件常伴随海量数据泄露、商业秘密侵权、网络平台运营秩序扰乱等多重法益侵害。一方面，通过网络爬虫非法获取数据的行为本身，可能对以数据为载体的特定法益构成直接侵害；另一方面，非法获取的数据还可能被用于电信网络诈骗、非法放贷、传播淫秽物品等下游犯罪活动，进一步放大法益侵害后果。基于此，为填补前置法规制漏洞、防范数据安全风险，亟需刑法发挥其在法律体系中的兜底保障作用。

2. 我国网络爬取数据刑法规制困境

我国网络数据爬取行为规制困境包括罪与非罪边界模糊、此罪与彼罪罪名适用混乱等问题。在规制爬虫行为的各规范之间衔接不明，导致立法落空，许多相关概念诸如“违反国家规定”“侵入”等决定入罪问题难以解释或解释标准差异较大，在司法实践中产生同案不同判等问题，损害司法公信力与法律适用的可预期性。

(一) 罪与非罪边界模糊

1) 前置法与刑法衔接断裂

纵观我国《刑法》，无论是第 285 条规定的非法获取计算机信息系统数据罪，还是第 253 条之一规定的侵犯公民个人信息罪，均将“违反国家规定”设定为成立犯罪的法定构成要件。根据《刑法》第 96 条的立法解释，“违反国家规定”主要涵摄两类规范性文件：一是全国人民代表大会及其常务委员会制定的法律；二是国务院制定的行政法规、发布的决定与命令，而部门规章、地方性法规等则未被纳入该范畴。虽有前置性规范对爬虫行为作出原则性约束，例如行政法规《网络数据安全管理条例》第 18 条确立了爬虫行为的基本合规框架[3]，以及 2025 年修订的《反不正当竞争法》亦新增数据竞争条款以回应数据抓取中的竞争秩序问题[4]，但此类规范均未对“合法爬取”的核心要素作出明确界定：诸如合理爬取频率的量化标准等关键内容仍处于模糊状态。这些疏漏直接导致《刑法》中“违反国家规定”的认定缺乏明确的规范依据，司法实践中对爬虫行为是否满足该要件常陷入无据可依的困境。

从规范功能来看，由于在立法中难以确认“违反国家规定”这一要件的认定标准，因此其在司法适用中更多仅具备形式认定意义，难以承载对爬取行为实质违法性的评价功能。因此，通过认定爬取行为对法益的侵害符合刑法评价逻辑。从网络爬虫的技术属性可能对被爬取计算机信息系统安全构成潜在威胁的客观事实出发，可明确：涉爬虫行为的非法性本质，在于对被爬取方计算机信息系统安全这一法益的侵害，其实质构成对计算机信息系统的“非法访问”。而判断某一爬取行为是否属于“非法访问”，核心标准在于该行为是否获得被爬取网站或平台的有效同意与明确授权，若缺乏授权或超越授权范围，即可初步认定其具有违反“国家规定”的潜在可能性[2]。

2) “侵入”概念的解释困境

我国现行刑法对网络数据爬取行为的规制，采用“专门罪名主导、兜底条款补充”的双层体系。其中，《刑法》第 285 条第 2 款规定的“非法获取计算机信息系统数据罪”，是规制此类行为的核心罪名。根据条文语义与司法解释精神，该罪的适用需满足“突破技术防护措施”与“获取计算机信息系统中存

[3]《网络数据安全管理条例》第十八条网络数据处理者使用自动化工具访问、收集网络数据，应当评估对网络服务带来的影响，不得非法侵入他人网络，不得干扰网络服务正常运行。

[4]此次修订新增 8 项条款，修改 16 项条款，进一步细化完善了不正当竞争行为的表现形式，并将非法获取数据、内卷式竞争、经营者滥用优势地位等新型不正当竞争行为纳入规制范围。

储、处理或传输的数据”两大要件，对于“侵入”概念的不同定义将极大程度影响网络数据爬取行为能否被认定构成此项罪名。

关于爬虫行为是否构成刑法意义上“侵入”的认定标准，学界与实务界形成两种观点，二者的核心争议聚焦于技术突破是否等同于刑法上的侵入。一种观点主张，“未经授权访问”的认定应以是否破解加密算法或核心安全防护措施为唯一标准，仅违反非强制性使用协议的行为，因未触及计算机信息系统的安全核心，不应纳入刑法评价范畴[3]。另一种观点主张技术层面的突破不等于刑法意义上的“侵入”。其核心论据在于：刑法的规制对象是具有实质法益侵害性的行为。随着数据表征权利客体的多样化，网络爬虫未经授权或超越授权抓取数据行为，依据被抓取数据所表征的不同法益，可构成不同罪名[4]。反爬措施的设置具有较强任意性，实践中可能存在数据持有方为维护商业垄断地位、过度限制数据流通而设置冗余反爬措施的情形，若仅突破此类措施而未对数据持有方的合法权益或公共利益造成实际侵害，不应认定为刑法上的“侵入”。反爬虫措施并不涉及访问权限，其作用只是限制访问的方式，故规避反爬虫措施的行为并不能评价为刑法意义上的侵入[5]。

(二) 此罪与彼罪的罪名适用混乱

1) 计算机犯罪与个人信息犯罪的竞合处理

数字经济时代，网络爬虫技术作为数据流转的重要工具，其合法性边界在实务中持续引发争议。我国《刑法》第 285 条、第 286 条规定的非法获取计算机信息系统数据罪、破坏计算机信息系统罪等罪名，与第 253 条之一的侵犯公民个人信息罪，在规制行为时呈现出法益交叉与构成要件重叠的特征。当爬虫同时爬取计算机系统数据与公民个人信息时，存在罪名适用冲突。部分法院对爬取含个人信息的数据直接认定为侵犯公民个人信息罪，忽略了计算机信息系统犯罪的适用条件；另有法院以“数据存储于计算机系统”为由优先适用非法获取数据罪，导致同案异判。例如，2018 年 9 月起，上诉人欧阳航宇、欧阳通在网上下载的专门软件，获取大量 QQ 号和 QQ 邮箱账号，并对其进行筛选、碰撞解密，再猜密保答案。广东省珠海市中级人民法院对上诉人欧阳航宇以非法获取计算机信息系统数据罪论处[5]。2023 年 7 月至 2023 年 10 月，被告人郭某在担任职务期间，为了维护客户关系，配合客户某某公司 1 经理任某在工作电脑中安装“爬虫软件”以获取德邦物流公司 UAP 系统中的快递面单信息。上海市青浦区人民法院认定其行为构成侵犯公民个人信息罪[6]。相似的社交平台数据爬取案件，有的认定为非法获取数据罪，有的则认定为侵犯公民个人信息罪。

2) 提供程序罪与共同犯罪的界分

网络爬虫技术兼具工具性与技术性双重属性，这使得围绕爬虫程序、接口及技术支持的提供行为，在刑事法律评价层面面临双重困境：一方面，该类行为可能落入《刑法》第 285 条第 3 款规定的提供侵入、非法控制计算机信息系统程序、工具罪(以下简称“提供程序罪”)规制范畴；另一方面，亦存在构成非法获取计算机信息系统数据罪、侵犯公民个人信息罪等下游犯罪共同犯罪的刑事风险。例如，2023 年 2 月起，被告人李某、吴某由于需要大量数据，二人便雇用了从事技术工作的被告人陈某编写非法爬取数据的接口，企图非法爬取相关数据。上海市普陀区人民法院认为陈某仅提供专门用于规避系统验证机制的有偿“爬虫”程序及接口，其行为符合提供侵入计算机信息系统程序罪的构成要件[7]。2015 年 1 月左右，被告人叶源星编写了用于批量登陆某宝账号、密码的“小黄伞”软件供他人使用，被告人叶源星将图片验证码识别(俗称“打码”)的业务交由被告人张剑秋协助完成。杭州市余杭区人民法院认为被告人叶

[5]参见广东省珠海市中级人民法院(2020)粤 04 刑终 101 号刑事判决书。
[6]参见上海市青浦区人民法院(2024)沪 0118 刑初 669 号刑事判决书。
[7]为生成商业报告，三人爬取 8 亿余条餐饮商超及地图数据获刑——今日头条 https://www.toutiao.com/article/7552787217232577050/?upstream_biz=doubao&source=m_redirect，2025/10/8。

源星、张剑秋结伙提供专门用于侵入计算机信息系统的程序，情节特别严重，其行为均已构成提供侵入计算机信息系统程序罪，且二人属于共同犯罪。相似的提供“爬虫软件”获取数据的行为，有的单独定罪，有的则构成共同犯罪。

3. 我国爬取数据行为刑法规制路径展开

针对上述有关爬取数据行为的刑法规制难点，本文认为应当构建“三阶层认定标准”以处理爬虫行为合法性认定问题，通过阶层化的审查框架明确爬取行为的刑事违法性判断逻辑；同时，以该阶层化框架为依托，对“违反国家规定”“侵入”等影响入罪与否的关键法律概念进行精准限缩，最终破解爬取行为入罪标准模糊、罪名适用错位等实践难题。

(一) 细化爬取数据行为的合法性认定标准

参照《网络数据安全管理条例》第 18 条，在司法解释中确立“三阶层认定标准”。综合两种观点，总结出以技术外观为第一审查指标，以是否产生实质侵害进行判断，最终以程序正当性审查作为权利保障的认定标准。

第一，以爬取数据技术作为前置判断标准，区分为禁止性技术、允许性技术以及功能性技术。禁止性技术是指具有避开或突破计算机信息系统安全保护措施，未经授权获取访问受限数据，且专门设计用于非法用途的爬虫技术。对爬取技术是否正当的性质认定，可以以网站设置安全保障机制的规范目的为核心基准，进行更为审慎的审查与分类界定。安全保障机制的本质属性，在于直接关联于系统防护、数据访问权限管控或风险抵御，其设立需服务于维护系统安全、保障数据完整性与防范非法访问的核心目标。例如，某理财 APP 的登录系统需“身份证号 + 人脸识别”，其目的是保护用户的理财资金数据，属于安全保障机制；而某视频 APP 的登录系统仅需“手机号验证”，登录后仍可免费观看公开剧集，其目的是收集用户手机号以推送广告，不属于安全保障机制，即使行为人绕过该登录系统抓取剧集信息，也不应认定为“侵入”。而禁止性技术的使用通常伴随着大量的并发请求和高频访问，破坏被爬取网站所设置的安全防护措施，最终超出了被爬取网站的技术承载能力，严重威胁被爬取网站的信息系统安全。例如超高频访问技术，该技术具有较强的主观恶性及网络危害性，表现结果就是网站负载过重而无法响应正常用户。其核心特征在于对目标网站正常功能的直接影响或破坏[6]。在三阶层判断中，此类技术直接满足构成要件该当性，无需考量数据内容即可推定违法性。这一观点在丁某提供爬虫软件案中得到印证[8]。允许性技术则是指对部分爬取数据行为认定合规的情况，此类行为模拟人工浏览、遵守速率限制且其未触发反爬机制，即使未能获得平台明示授权时，也不触及刑法防线。功能性技术兼具合法与非法功能，对其性质判断则需结合案情综合分析判断是否能够适用技术中立原则，进而排除其违法性。

第二，以法益是否受到实质侵害作为判断爬取数据行为是否具有实质违法性的基准，具体审查内容包括爬取数据中是否包含受刑法保护的个人信息、是否影响平台数据管理权与市场竞争秩序、是否侵犯著作权、是否侵犯商业秘密等，同时将比例原则嵌入实质违法性判断，综合审查实现法益平衡，既保护合法数据利用，又防范刑法过度干预。实质违法性是三阶层入罪判断的核心过滤层，指爬虫行为通过技术手段获取数据的过程中，对刑法所保护的法益造成现实、严重且值得刑罚处罚的侵害或威胁。其核心功能是排除形式符合构成要件但实质无害的行为，聚焦于法益侵害的实质内核，而非仅停留在技术手段的形式判断。以个人信息数据为例，关于爬取个人信息数据的入罪标准，首先以“结合可识别”对个人信息数据进行过滤。目前，许多互联网公司能够使用各种方式获取海量信息，并通过技术手段对其进行

[8]参见最高人民法院入库参考案例：丁某提供侵入计算机信息系统程序案(入库编号：2024-18-1-253-001)。本案中被告人丁某向他人提供的“客多多精准获客”软件专门用于入侵短视频平台服务器非法获取未授权人员访问受限的数据，属于刑法规定的“专门用于侵入、非法控制计算机信息系统的程序、工具”。

系统处理及分析运算，最终获取自然人的各类身份信息。因此，采用“单独识别 + 结合识别”双重标准是与时俱进的手段，能够将手机号、人脸信息等直接标识，以及设备指纹、UID 等可结合其他信息定位特定自然人的间接标识均纳入审查范畴。只有所爬取的个人信息数据是具有可识别性的，才能够进入刑法视野。其次，考量个人信息数据的重要程度。结合 2017 年《刑事司法解释》《个人信息解释》的规定进而区分出敏感信息、重要信息和一般信息。敏感个人信息数据一旦泄露或非法使用，容易使被收集者的人身、财产安全受到严重影响，因此对敏感个人信息必须征得被收集人的明示同意，并且采用“零容忍”标准，只要存在爬取行为且数量达到入罪阈值即触发刑事评价；对普通个人信息则需满足“数量较大且用途违法”双重条件。

与此同时，判断数据爬取行为的合法性也应当结合比例原则综合判断。比例原则作为实质违法性审查的核心分析工具，可以参照行政法领域适用比例原则的有益经验，一般可以概括为三个子项：适当性、必要性和衡量性递进审查，构建爬虫行为刑事评价的精细化框架[7]。其一，适当性即目的合法性审查是比例原则适用的前置前提。爬虫行为的目的需契合法律规定与公共利益导向，合法目的应限定为学术科研等公共利益维护或自身合法经营需求等具有正当性基础的范畴。若爬取目的指向非法牟利、滋扰他人、不正当竞争等违法性诉求，则径行推定其目的具有非正当性，进而强化对行为违法性的否定性评价，为后续刑事归责奠定基础。其二，必要性审查聚焦于行为方式的谦抑性要求。在目的合法的前提下，爬虫行为需选择对法益损害最小的实现路径，核心审查维度包括：是否存在更具谦抑性的替代路径；是否合理规制爬取频率与规模，避免因高频次、大规模请求导致服务器负载超标、响应迟缓甚至宕机等损害后果；是否恪守数据获取的必要性边界，仅爬取实现合法目的所需的最小范围数据，而非无差别全面抓取。唯有满足上述要求，方能认定其手段具有必要性，排除部分违法性评价。其三，衡量性审查旨在实现行为收益与损害后果的动态平衡。审查核心在于判断爬虫行为所产生的正向价值与对刑法保护法益造成的损害是否处于均衡状态。造成损害的人，不一定要负赔偿、缴税或禁止活动等法律责任，因为这是两个不相容活动之冲突，若是造成损害的活动价值高，则造成损害之人则有理由不承担不法责任。在权益相互冲突的环境下，数据爬取还是不爬取都是相互性的[8]。例如，短期便利消费者的爬虫行为，若长期侵蚀数据权利人的劳动成果、破坏行业创新生态，即因损害大于收益不符合相称性要求；反之，仅造成轻微数据泄露且行为人及时采取补救措施、未引发实质危害后果的，可认定为法益失衡，排除实质违法性。比例原则通过上述三阶审查，既实现对值得刑法处罚行为的精准锁定，又发挥对形式突破技术限制但实质无害行为的出罪功能，形成爬虫行为刑事评价的筛选机制。在李开祥案中，行为人采用欺骗性技术手段非法获取 8100 万余条包含敏感个人信息的公民个人信息，既对公民个人信息权益造成实质侵害，又因爬取目的非法、手段缺乏必要性与相称性，完全背离比例原则要求，最终被认定构成侵犯公民个人信息罪[9]。而“白帽子”爬虫行为中，行为人通过技术手段协助平台发现安全漏洞、完善防护机制、提升网络安全水平，其行为所产生的公共利益价值显著大于潜在法益损害，经相称性审查符合比例原则，得以排除实质违法性。综上，比例原则与法益侵害实质判断的结合，既避免了仅以技术形式为标准的机械归罪，又防止了刑事处罚范围的不当扩张，为爬虫行为的刑事评价提供了兼具合法性与合理性的分析框架，实现了数据利用自由与法益保护的动态平衡。

第三，以违法阻却事由与补救机制作为权利保障的兜底防线。内容具体包括平台授权、用户授权以及程序补救措施。平台授权需区分明示授权和默示授权。明示授权是平台通过书面、电子等可查证形式，对爬虫行为作出的针对性许可，包括时间边界、范围边界及手段边界。默示授权是平台未作明确表示，但通过行为或行业惯例可推定同意爬虫行为的情形。平台未采取密码验证、数据加密、Robots 协议禁止

[9]参见最高人民法院指导案例 192 号：李开祥侵犯公民个人信息刑事附带民事公益诉讼案。

等防护措施。默示授权的认定需要更审慎的证据支撑，以平衡数据利用与权利保护的冲突。这种严格化导向符合我国刑法谦抑原则对入罪边界的精准把控要求。例如，爬取方需举证平台未采取具有实质排除效力的防护措施，且该“不防护”是平台的主动选择而非技术疏漏。技术障碍的缺失是默示授权的前提，但需区分主动放弃防护与被动未防护，仅有前者可作为授权推定依据，避免将平台技术能力不足等同于授权意愿。爬取方还需举证爬取行为符合行业惯例，且依前文所述爬取的数据未超出“合理使用”范畴。美国第九巡回上诉法院在 hiQ Labs 诉领英案中明确，行业惯例行业不断发展，行业主体的共识也会逐渐发生变化[9]，行业惯例的认定需结合公共利益，仅当惯例不损害市场竞争秩序与他人权益时，方可作为默示授权的依据 [10]；我国学者在著作权默示许可研究中也强调，行业惯例需满足合法性前提，不得与法律强制性规定冲突。构建严格化的默示授权证据标准，本质是通过程序正义保障实体权利，既避免平台以未明示授权随意否定合法爬取行为，也防止爬取方滥用默示授权规避法律责任。正如学者在爬虫协议研究中所强调的，默示授权的价值在于补充明示授权的不足，而非成为规避授权义务的通道。从司法实践来看，这一标准与我国“百度诉奇虎 360 不正当竞争案”[11]“hiQ Labs 诉领英案”等典型判例的裁判逻辑一致，均强调默示授权的认定需建立在充分证据基础上，拒绝“推定优于证据”的宽松倾向。未来，随着数据分类分级保护制度的完善，对于涉及核心数据、敏感个人信息的爬取行为，默示授权的证据标准应进一步收紧，甚至原则上排除默示授权的适用。用户授权则是数据主体对权利的让渡。数据爬取的合法性必须满足权利主体对授权行为及其内容的充分知情，爬取敏感个人信息必须取得用户单独书面授权，概括授权不具效力。另外，关于公开数据的爬取问题，若公开数据属于“结合可识别”范畴，则仍需用户授权。基于刑法规制爬取公开数据行为应坚持的刑法谦抑立场，在数据的信息内容已经公开的情况下，不宜以保护数据载体保密性为由广泛适用非法获取计算机信息系统数据罪规制爬取公开数据的行为，而只能在爬取的数据属于刑法已经类型化保护的重要数据时，适用相应罪名进行规制爬取行为[10]。

(二) 明确相关概念

“违反国家有关规定”应当属于空白罪状而非叙明罪状。当刑法条文没有对构成特定罪名的具体行为要素作出明确具体的规定时，就需要借助有关法律法规或部门规章作出具象且清晰明确的规定[11]。有关“国家有关规定”的概念，需明确其范围仅涵盖法律、行政法规及经法律授权的部门规章，目前立法中没有明确规定爬取频率的量化标准。不过，《数据安全管理办法(征求意见稿)》第 16 条首次提出“自动化访问收集流量超过网站日均流量三分之一”等量化方向，为后续司法实践中细化具体数值提供了立法雏形。实践中，法院往往将“三分之一流量”细化为“5 倍流量”，将“高频访问”具象为“1 秒间隔且持续 1 小时”，形成可操作的裁量基准。2023 年 5 月，被告杨某使用 A 公司临时授权账号，通过爬虫软件在 48 小时内发起高频访问 6 万余次，单 IP 峰值请求间隔低于 0.1 秒，累计爬取数据 1800 余万条，远超授权的“单日查询 500 条、单次导出 100 条”限制，导致 A 公司 P 系统负载率飙升至 92%，被迫关停 48 小时。该判决未直接援引“5 倍流量”指标，但通过“授权范围对比”间接量化：实际爬取量为授

[10]hiQ Labs (以下简称“原告”)就雇员职业发展角度，为雇主提供基于数据的咨询服务；LinkedIn (以下简称“被告”)运营全球最大的职场社交网络，有超过 5 亿用户。原告服务依赖于对被告数据的抓取。被告对自身网站爬虫协议有严格限制，仅限特定主体抓取。被告允许用户采取灵活的隐私设置，亦采取爬虫识别系统、黑名单等多种手段限制数据爬取。2017 年，大致与被告公开宣称计划推出与原告类似的服务同时，被告向原告发去警告函，要求原告停止爬取被告网站数据。原告遂于加州北区联邦法院以侵权妨碍等多种诉由起诉，并向法院申请诉中禁令，要求禁制被告限制其爬取数据各类行为。法院支持原告请求，颁布诉中禁令，被告遂上诉至联邦第九巡回法院。上诉法院维持禁令。

[11]2013 年 10 月 16 日，此案在北京市第一中级人民法院公开开庭审理。百度公司当庭表示，奇虎公司在经营 360 搜索引擎的过程中存在对百度的不正当竞争行为，具体行为包括：一、违反搜索引擎的机器人协议(Robots 协议)，擅自抓取、复制原告网站并生成快照向用户提供；二、在原告明确函告被告后，仍擅自抓取、复制原告网站并生成快照向用户提供；三、绕过原告网站，在用户点击搜索结果中原告的网站地址后，直接向用户提供快照服务。2014 年 8 月 7 日，围绕 360 搜索引擎是否违反 Robots 协议而引发的不正当竞争纠纷一案，北京一中院作出一审判决，认为被告奇虎公司的行为违反了《反不正当竞争法》相关规定，应赔偿原告百度公司经济损失及合理支出共计 70 万元，同时驳回百度公司的其他诉讼请求。

权上限的 36,000 倍，日均请求量占网站同期正常流量的 8.7 倍，远超《数据安全管理办法(征求意见稿)》“三分之一流量”的立法雏形 [12]。但是，该文件至今未正式生效，不属于严格意义上的“法律、行政法规”，因此可以将该指标纳入立法考量范围，明确“妨碍网站运行”的频率量化参考。

有关“侵入”行为的界定，基于前文对两种理论的分歧剖析，以及爬取数据行为入罪的“三阶层认定标准”研究，本文主张应遵循形式违法性与实质违法性相统一的刑法原理，构建二元融合认定路径。该路径的核心在于，将技术标准作为“侵入”行为的形式判断基准，将法益侵害作为“侵入”行为的实质评价要素，以刑法谦抑性保障评价结果的合理性。侵入的本质特征就是实质上的未经授权的访问。数据的访问权限是界定网络爬虫法律责任的关键要素，要求对抓取是否被“授权”进行规范性解释[12]。上海晟品公司爬虫入罪案被视为我国司法实践中爬虫行为从民事领域转向刑事领域的里程碑案例。该案中，晟品公司的主管人员为牟取非法经济利益，组织技术人员开发专用爬虫程序，通过突破原告企业设置的数据安全防护屏障，破解身份验证机制、绕开访问频率限制，批量抓取原告平台存储的海量正版视频数据。法院经审理认为，晟品公司的行为已超出民事侵权范畴，其突破安全防护屏障、非法获取计算机信息系统中存储的数据的行为，符合《刑法》第 285 条第 2 款非法获取计算机信息系统数据罪的构成要件，最终对相关责任人员判处刑罚。此案的裁判逻辑明确了爬虫行为构成刑事犯罪的核心标准：一是技术手段具有“非法性”，即突破安全保护措施；二是获取的数据属于计算机信息系统中存储的具有价值的数据；三是行为造成了严重后果，爬取数据量巨大、权利人损失惨重[13]。在数字经济高速演进的时代语境下，计算机类犯罪中“侵入”行为的界定标准不宜采取静态固化模式，而应秉持动态调适立场。此种调适需求的核心逻辑在于：一方面，数据保护制度正从零散规范向体系化架构逐步完善，《数据安全法》《个人信息保护法》等法律的实施细则与司法解释持续细化，对数据法益的分层保护体系不断健全；另一方面，爬虫技术呈现迭代升级态势，技术手段的复杂性与多样性对“侵入”的形式判断提出新挑战，爬虫技术本质上有利于信息的交流和共享，因此并不必然受到禁止或需要法律予以规制[14]。基于此，“侵入”界定需伴随制度完善与技术发展，持续细化技术要件的层级划分标准，并优化法益侵害的认定基准。最终，为数据安全保障与数据合理利用之间的价值平衡筑牢刑法层面的制度支撑，既防止技术滥用对数据法益造成的实质侵害，又为数字经济创新发展保留必要的制度空间。

参考文献

[1] 陈毅坚, 曾宪哲. 网络爬虫刑法规制研究[J]. 广东社会科学, 2022(5): 240-253.

[2] 孙杰. 数据爬取的刑法规制[J]. 政法论丛, 2021(3): 115-125.

[3] Kerr, O.S. (2003) Cybercrime’s Scope: Interpreting “Access” and “Authorization” in Computer Misuse Statutes. *New York University Law Review*, **78**, 1596-1668.

[4] 刘艳红, 杨志琼. 网络爬虫的入罪标准与路径研究[J]. 人民检察, 2020(15): 26-31.

[5] 孙禹. 强行爬取公开数据构成犯罪吗[J]. 国家检察官学院学报, 2021, 29(6): 121-139.

[6] 刘云. 网络数据爬取合法性判定的三阶层认定标准[J]. 东方法学, 2025(4): 37-38.

[7] 李雷. 论数字时代个人信息保护与利用平衡的展开路径[J]. 行政法学研究, 2024(1): 111-122.

[8] 许可. 数据爬取的正当性及其边界[J]. 中国法学, 2021(2): 166-188.

[9] 曹丽萍. 爬虫协议作为商业道德评判行为正当性的考量维度——评北京字节跳动科技有限公司与北京微梦创科网络技术有限公司不正当竞争纠纷案[J]. 法律适用, 2023(5): 95-104.

[10] 石经海, 苏桑妮. 爬取公开数据行为的刑法规制误区与匡正——从全国首例“爬虫”入刑案切入[J]. 北京理工大学学报(社会科学版), 2021, 23(4): 154-164.

[11] 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.

[12]合法授权不等于无限授权_中华人民共和国最高人民检察院 https://www.spp.gov.cn//zdgz/202507/t20250715_701316.shtml，2025/10/10。

[12] Goldfoot, J. and Bamzai, A. (2016) A Trespass Framework for the Crime of Hacking. *The George Washington Law Review*, **84**, 1477-1499.

[13] 阮林赟. 网络爬虫刑事违法的立场、标准和限制载[J]. 河北法学, 2021(7): 173-187.

[14] 林维. 数据爬取行为的刑事司法认定[J]. 人民检察, 2020(4): 38-40.