

“网络爬虫”法律风险与合规路径

——基于 242 份司法裁判文书的实证研究

张 泳¹ 闫芊伊²

摘 要：“善意爬虫”促进商业模式创新与数据要素市场建立的同时，“恶意爬虫”造成了公民、企业、政府信息数据的安全隐患。如今，网络数据爬取行为诉诸刑事规制已成为司法实践事实，且呈现逐步扩大化的趋势。然而，既往有关“网络爬虫”法律规制的研究多为理论分析，缺少实证层面的验证，且未能以“全景”的观察范式绘制出其治理谱系。在此背景下，通过对 242 份司法裁判文书进行统计分析，运用实证分析的研究方法，分别梳理了“网络爬虫”行为本身及后续行为的刑事犯罪、行政违法与民事侵权法律风险。并根据爬取数据的类型，为个人、企业及政府数据的“网络爬虫”行为分别制定了合规路径，希冀明晰“网络爬虫”全局治理的合规路径，为有关专家、学者及实务界同仁提供有益参考。

关键词：网络爬虫；刑事犯罪；民事侵权；行政违法；合规路径

一、问题的提出

在数字经济时代，各个国家、各行各业都高度认可数据的经济要素地位，更是引起了党中央、国务院的高度重视。在党的十九届四中全会上，数据要素的地位第一次得到了认可，随后，在十九届五中全会上，数据要素的市场地位被再次强调。2022 年 10 月 16 日，党的二十大会议更是提出了“加快建设数字中国”的战略目标。数据在带来价值张力的同时，亦诱发刑事犯罪，其中，“网络爬虫”爬取数据的行为极具代表性。³ 作为一项新兴信息技术，“网络爬虫”可以用于处理新兴技术开发的网页、满足日益增多的信息需求及爬取有用

1 张泳（1994-05），男，珠海市横琴粤澳深度合作区研究院特聘研究员，澳门大学法学院博士研究生。联系电话：+8618827872625；邮箱：2507428407@qq.com。

2 闫芊伊（1997-11），女，澳门大学科技学院硕士研究生。联系电话：+8613203452922；邮箱：2439070747@qq.com。

3 “网络爬虫”是指通过爬虫软件自动抓取第三方数据库数据，表现为“网页抓取”（web scraping）、“屏幕抓取”（screen scraping）等爬取形式。See Jennings, F. & Yates J., Scrapping Over Data: Are the Data Scrapers' Days Numbered?, Journal of Intellectual Property Law & Practice, 2009, Vol.4, pp.120-129.

的关键信息等。网络爬虫因其技术的中立性，本身并没有善恶之分，但在数据价值的驱动下，大量任意爬取数据、无视规则的行为不断发生。因此，在理论上，网络爬虫被分为“恶意爬虫”与“善意爬虫”。⁴其中，“恶意爬虫”正吞噬着正义价值，侵害他人权益、违背商业道德、消解人格尊严、打破平等自由，渐次生成了数字化社会的“算法蛀虫”。⁵例如，大量恶意爬虫爬取公民、企业及政府等的信息，对公民的个人信息、网络信息安全、国家安全等造成的侵害；再如，恶意爬虫行为通过毁坏计算机信息系统，同时散播计算机病毒，将对国家网络系统造成威胁，甚至是造成国家秘密的泄露；又如，同时对某一网站使用大量网络爬虫，致使网站的崩溃，并对网络的稳定性造成影响。为了调整网络爬虫技术使用的无序状态，必须及时规范并调整“恶意爬虫”行为。而司法作为社会公正的最后一道防线，从司法裁判中分析网络爬虫行为刑事犯罪的特点与困境，将为立法工作者、司法实务人员及相关领域的学者提供有益参考。

从研究现状看，目前学界对于网络爬虫的学术研究主要集中在理论研究层面，实证研究较少。有学者从抓取对象入手，通过阿历克西“权重公式”对爬取企业数据、政府数据、个人数据的正当性边界进行划定⁶；有学者从竞争法层面展开研究，认为应该“促进以‘保护与竞争’为基调的数据爬取行为竞争法的治理理念、规则、方式”⁷；也有学者从刑法入手，认为“抓取开放数据和单纯违反行业规则的爬虫行为无需适用刑法，突破反爬防护措施和抓取非开放数据并非都要承担刑事责任”，具体而言，有的涉及“非法获取计算机信息系统数据罪”的分析⁸，有的涉及“侵犯公民个人信息犯罪”的探讨⁹。

总体而言，学界对网络爬虫的法律规制进行了许多有益的理论贡献，但相关的理论、原则、方法对法律实践的助益，仍需要实证研究的检视。因此，从司法裁判文书中找寻实践规律，促进相关理论研究的有效转化，是纵深规制网络爬虫行为的应有之义。基于此，本文以“北大法宝”为检索平台，在“司法案例”栏目中，全文搜索“爬虫”或者“爬取”，审结日期设置在“2020年1月1日”至“2022年12月31日”，案由为“刑事”“民事”，并以“判决书”为文书类型进行高级检索，获得民事判决书290份、刑事判决书63份。经

4 参见孙立伟、何国辉、吴礼发：《网络爬虫技术的研究》，载《电脑知识与技术》2010年第15期。

5 正义包括“禁止剥夺他人利益、社会服务、人格尊严与自由”三大内涵以及“平等自由、公平与差异”两大原则。参见[美]约翰·罗尔斯：《正义论》，何怀宏、何包钢、廖申白译，中国社会科学出版社2009版，第60-516页。

6 参见许可：《数据爬取的正当性及其边界》，载《中国法学》2021年第2期。

7 参见陈兵：《保护与竞争：治理数据爬取行为的竞争法功能实现》，载《政法论坛》2021年第6期。

8 参见郭玮：《网络数据爬取行为的刑法规制研究——以非法获取计算机信息系统数据罪为视角》，载《新疆社会科学》2020年第3期。

9 参见刘艳红：《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》，载《政治与法律》2019年第11期。

过筛选，剔除重复或者与“网络爬虫”无关的案例后，共获得民事判决书 174 份、刑事判决书 61 份。¹⁰ 而爬虫行为的行政法规制在行政诉讼中并没有体现，司法实践中，爬虫行为主要受到行政处罚的规制。因此，在“北大法宝”平台中，将上述除“文书类型”外的关键词在“行政处罚”一栏检索，并在“威科先行”平台进行验证。最终，检索得到行政处罚决定书 21 份，经剔除后获得符合要求的处罚决定书 7 份。综上，本小组共检索到相关案例 374 个，筛选掉不符合研究目的的案例后，共获得有效样本 242 份。下文将以相关司法裁判为研究基础，从多个维度分析网络爬虫行为所涉的法律风险，并以裁判规则为导向提出行为合规的建议。

二、“网络爬虫”的法律风险分析

网络爬虫技术的不当使用，存在多种不同的法律风险。根据行为类型进行划分，存在使用爬虫技术行为本身涉及的法律风险，和对所获取数据进行处理等后续的数据使用行为。通过对全国涉及违法使用爬虫技术的判决书及行政处罚决定书等统计发现，54.8% 的案件仅针对数据获取后的使用、传播等行为，14.5% 的案件仅针对爬虫技术使用行为本身，而 24.2% 的案件存在着前述两种行为，另有 6.5% 的案件针对上述两种行为之外事实，¹¹ 见图 1¹²。

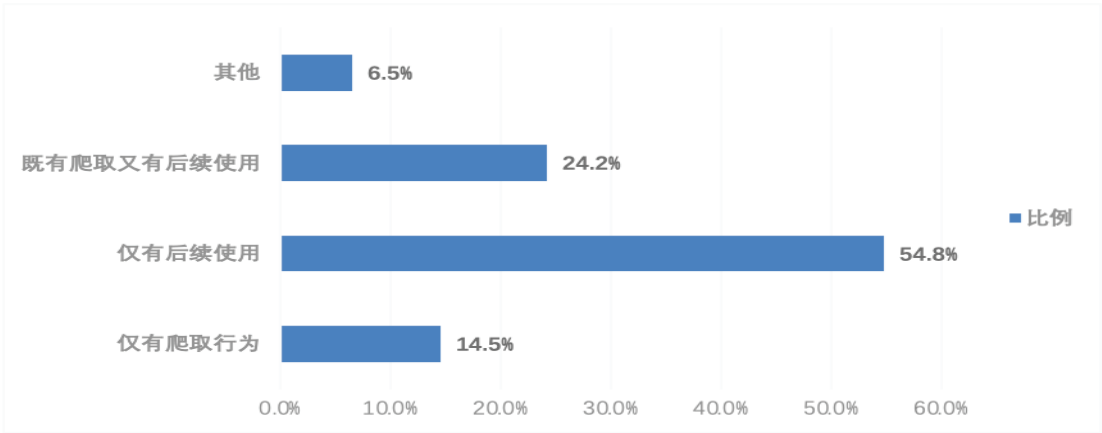


图 1：“网络爬虫”法律风险的分布

10 需要注意的是，由于北大法宝平台在案由部分将“知识产权与竞争纠纷”单独列出，但该部分案件在法信等平台中被归纳为民事案件，为了统计方便，我们将“知识产权与竞争纠纷”以及“民事”案件共同归为民事案件。

11 例如，某公司因未依法落实计算机信息系统安全等级保护制度，致使其管理的网站被恶意程序控制而受到了行政处罚。

12 本文图表均是对 242 份有效样本统计分析、整理后绘制而成。

（一）内部风险：基于爬取行为自身的法律风险

个人信息是数字时代一切财富与经济的根本出发点，因为一切被数字化记载于网络空间的个人信息都具有商业价值，例如产品开发、消费画像及精准营销等。而网络爬虫为海量信息的分类获取提供技术基础，“某种意义上讲，只要涉及信息搜集，就势必会运用到爬虫技术”，¹³但在当前隐私保护相对薄弱、数据安全漏洞频出的网络环境下，爬虫技术因其与信息来源违法、过度爬取等问题，由原本的“中立技术”逐步向“道德可疑”直至“违法技术”嬗变。在司法实践过程中，根据使用爬虫技术行为可能侵害的主体法益，可以将该行为法律风险归为四类，即涉嫌民事侵权、不正当竞争行为、行政处罚及刑事犯罪。据相关案例统计数据显示，因运用爬虫技术行为本身被判断为违法的案件中，刑事犯罪类的占44.4%，不正当竞争类的占33.4%，民事侵权类的占22.2%，见图2。而在既有针对爬虫行为本身又有针对后续数据使用行为的案件中，仅有被行政处罚及刑事犯罪类的，分别占比86.7%、13.3%。

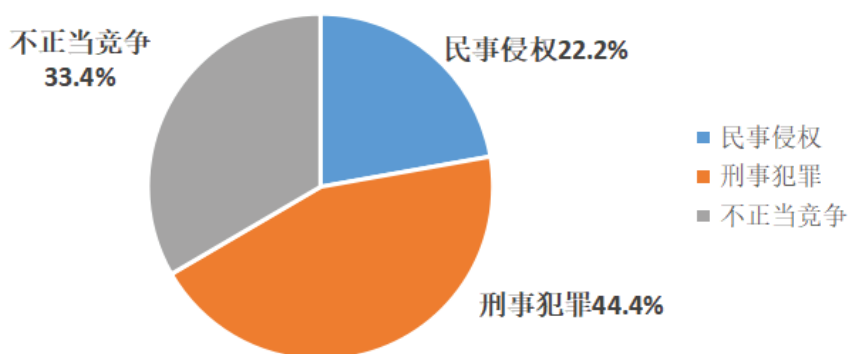


图2：因爬虫行为本身被判断违法的案件中违法类型分布

1. 涉嫌民事侵权

近年来，对于公民个人信息保护的呼声日益高涨，引起了立法的重视。比如《个人信息保护法》规定了个人信息主体的查询、复制、撤回同意、请求删除等权利，以及承担损害赔偿等侵权责任等责任形式。再如，《民法典》规定了停止侵害、消除影响等责任形式。又如，《数据安全法》第三十二条规定了“任何组织、个人收集数据，应当采取合法、正当的方式，不得窃取或者以其他非法方式获取数据”。在司法实践过程中，据上述民事案件统计数据分析，94%是因对所获取数据的提供、传播等后续的数据使用行为侵权，而因爬虫技术使用行为本身侵权的仅占6%，见图3。

13 参见林维：《数据爬取行为的刑事司法认定》，载《人民检察》2020年第4期。

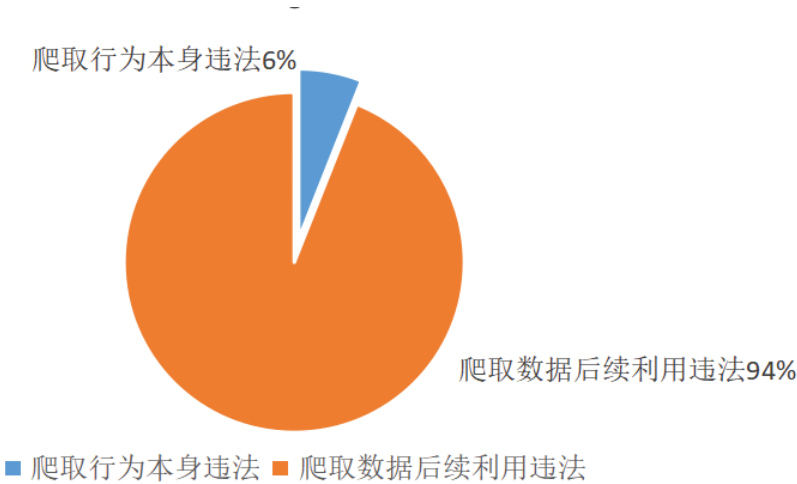


图 3：民事案件爬取行为类型

并且，在因爬虫技术使用行为本身侵权的案件中，爬取对象为公民的为 57.31%，爬取对象为企业的占了 39.39%，爬取对象为政府机关的仅占 3.3%，见图 4。

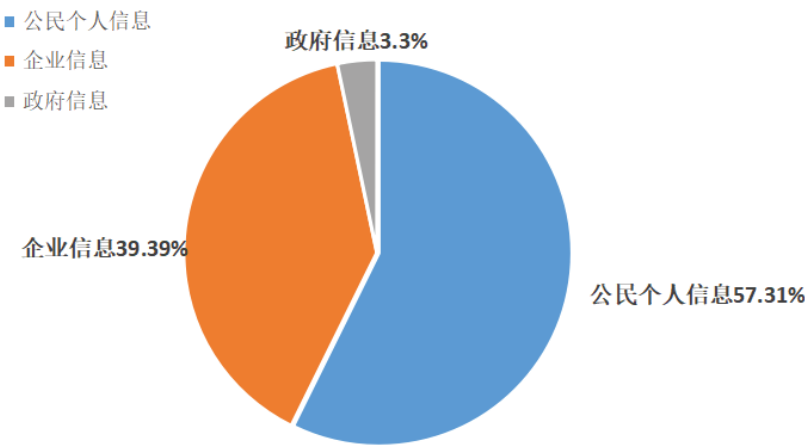


图 4：民事案件中爬取各类对象的占比

按照数据主体的划分，可以将使用爬虫技术行为的民事法律风险分为以下两类：

（1）侵害公民个人信息权益。依据《个人信息保护法》《民法典》，当前个人信息保护规范的要旨是个人信息收集和使用应遵循合法、正当、必要等原则，经被收集者同意收集和保存后应当依法处理或使用。因此，网络抓取公民个人信息行为是否合法的根本在于是否获得了信息主体的真正授权同意。一方面，个人同意并不完全等于授权，根据《个人信息保护法》第十三至十六条、第三十一条和《民法典》第一千零三十五条，“征得该自然人或者其监护人同意”及“在同意的范围内合理实施”均为个人信息处理行为合法的条件。可见，

主体同意并没有为信息处理者创设自由，不产生“授权”的法律效果。另一方面，司法实践过程中，个人信息的合法存在依赖于具体的场景。常见的情形是，虽然用户已将个人信息发布于具体的平台上，但第三方平台仍不得基于未经授权的爬取而随意使用，否则将构成对公民个人信息的侵权。例如，在孙某某诉北京某科技一案中，某公司通过搜索引擎爬虫技术从校友录网站上抓取原告孙某宝带有其面目特征信息的头像照片，并通过关键词搜索加结果展示的形式将“孙某某”这一自然人姓名和照片进行关联，使得个人信息脱离了“校友录网站上传的个人证件照和真实姓名仅限班级同学可见”的具体场景，且在孙某某两次通知后仍未采取措施消除不良影响，被法院认定为侵犯了孙某某的个人信息权益。¹⁴

（2）侵害企业数据财产性利益。信息时代，许多网络运营商掌握了大量的公民个人信息，甚至在某些领域存在着垄断相关数据的情况。不仅是公民个人信息，还有其他方面的海量数据，这些数据对于企业在某一领域的竞争力来说都是至关重要的。企业不仅仅是通过一定的方式获得了公民的个人信息或者其他信息，还会通过对数据的分析、加工等方式，来实现利用数据盈利的目的。通过分析、加工的这些信息就成为了企业独有的数据，而对于这部分具有独创性的企业数据如果利用网络爬虫技术非法获取，就可能侵害企业的知识产权或商业秘密。¹⁵ 在当前的司法实践中，企业信息的民事权利的保护方面，较为常见的是通过著作权来对滥用“网络爬虫”技术行为进行民法规制。例如，中文在线数字出版集团有限公司诉深圳聚领威锋科技股份有限公司案中，深圳聚领威锋公司在其应用程序“石头阅读特别版”APP上显示：“实时追踪网络小说更新，并第一时间提醒，我们使用最新的网络爬虫技术，智能帮你追踪你喜欢的网络小说，作者更新后，我们及时搜索整个中文网络，并整理好格式第一时间呈现给你……”，被告未经许可在该程序中向公众提供了涉案作品的下载服务，侵害原告享有的信息网络传播权及获得报酬的权利，被法院认定为侵害原告作品网络信息传播权，必须停止侵害、赔偿损失。¹⁶

14 北京市互联网法院（2019）京 0491 民初 10989 号。

15 例如，当爬取的用户数据具有独创性能够成为著作权的客体时，《著作权法》可以对其进行有效规制。《著作权法实施条例》第二条规定了作品的含义：“文学、艺术和科学领域内具有独创性并能以某种有形形式复制的智力成果。”参见《中华人民共和国著作权法实施条例》，http://www.gov.cn/zhengce/2020-12/26/content_5573623.htm，最后访问时间：2022 年 8 月 18 日。

16 北京市东城区人民法院（2018）京 0101 民初 15755 号。

2. 涉嫌不正当竞争

在《反不正当竞争法》的视域下审视“网络爬虫”技术的运用，主要聚焦于运用数据可能会造成扰乱市场竞争秩序，破坏市场公平机制等方面。对数据竞争秩序的维护包括对公平数据竞争秩序的维护和对自由数据竞争秩序的维护。¹⁷ 对此，司法实践中，主要通过《反不正当竞争法》第二条原则性规定及第九条有关商业秘密的规定进行裁决。即如果爬取的数据是《反不正当竞争法》第九条规定的具有秘密性、价值性、保密性的非公开用户数据，构成侵犯商业秘密的，可以直接通过《反不正当竞争法》第九条进行规制。如果被爬取的数据不属于商业秘密，因为利用“网络爬虫”技术这一方式并没有具体规定在《反不正当竞争法》中的不正当竞争类型中，因此，法院对该行为的判断是通过《反不正当竞争法》第二条这一原则性的条款，来积极地弥补了这方面的漏洞。在爬虫不正当竞争案例中，57.1% 是依据《反不正当竞争法》第二条作出的裁判，28.6% 是依据《反不正当竞争法》第九条，依据其他法条作出裁判的占比 14.3%，见图 5。

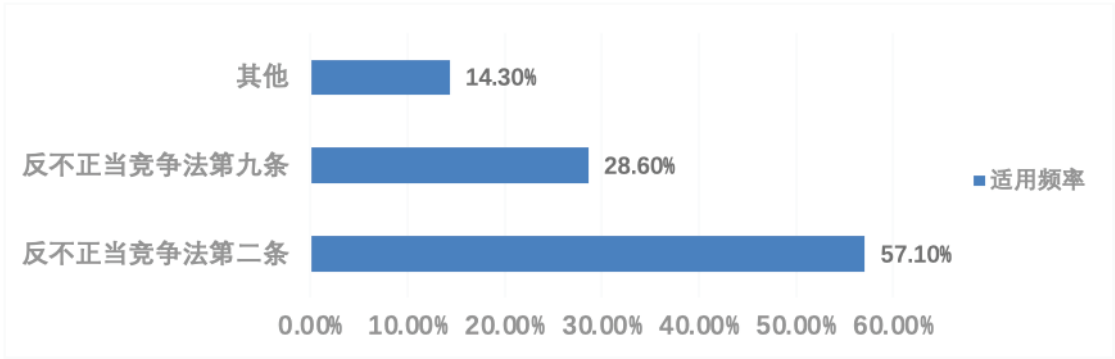


图 5：反不正当竞争法适用情况

而实践中，主要有两类不正当竞争行为，即无视网站设置的 robots 协议而随意抓取网站内容行为和绕过或破坏反爬措施¹⁸ 爬取数据的行为，分别占比 38.1%、32.3%，其他行为占比 29.6%，见图 6。

17 所谓公平数据竞争，是指应当警惕市场主体利用“网络爬虫”技术，违反 robots 协议，恶意使用技术，盗取其他合法拥有数据资源的市场主体的数据，损害对方权益而使自己的市场份额轻易增加的情形。所谓自由数据竞争，是指应当数据的自由竞争，避免某些市场主体过度占有某些数据而造成自由竞争的某一领域或行业的市场垄断的情形。参见陈军标、杨兰：《网络爬虫技术的法律规制》，载《上海法学研究》集刊 2020 年第 12 卷。

18 反爬措施是指利用某种技术手段阻止他人利用爬虫技术大规模访问自己客户端的方式，“IP 访问量限制、session 访问量限制、User-Agent 限制以及设置登陆验证码”都属于实践中常见的反爬措施。参见唐松：《Python 网络爬虫从入门到实践》（第 2 版），机械工业出版社 2019 年版，第 1 页。

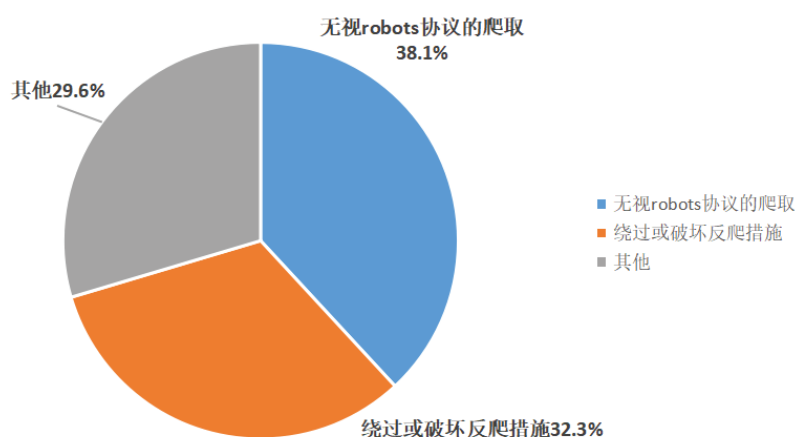


图 6：不正当竞争行为类型

（1）无视网站设置的 robots 协议而随意抓取网站内容行为可能违反《反不正当竞争法》构成不正当竞争。爬虫协议又称为机器人协议 (robots 协议)，是互联网站所有者基于商业竞争、数据安全或用户隐私，利用 robots.txt 文件向网络机器人 (Web robots) 给出网站指令的协议。作为一种“代码规制”，爬虫协议从技术上并不能阻止目标数据被爬，但目前已得到互联网搜索行业的普遍遵循，一些法院判决也认可其效力。¹⁹ 国内多家互联网搜索引擎服务公司还通过行业协会约定，开展搜索引擎服务要“遵循国际通行的行业惯例与商业规则，遵守机器人协议”、限制搜索应当具有“行业公认合理的正当理由”、不得利用协议“进行不正当竞争行为”。²⁰

（2）绕过或破坏反爬措施爬取数据同样有可能违反《反不正当竞争法》之规定²¹。自爬虫技术诞生以来，网络空间爬虫与反爬虫的争议和斗争就从未停止。数据权益语境下，以特定数据为目标，无视平台的多重反爬技术设置，利用爬虫技术与模拟登陆、秒拨动态 IP 等技术的组合升级来强行抓取数据的行为具有明显的非法性。²² 以基于 robots 协议禁止网络爬虫抓取信息的新浪微博为例，虽然定期升级自身的反爬措施，但还是会经常会面临一些数据收集者，通过建立大量微博账户来模拟正常用户行为，或者购买大量 IP 来伪造调用 IP 来源，通过伪造为正常用户的请求在互联网主页、移动客户端等进行数据抓取。因此，如果爬虫使用者未经授权利用爬虫技术采集了竞争对手的数据用于自身企业的经营，该“搭

19 北京市知识产权法院 (2017) 京民终某号判决书。

20 中国互联网协会：《互联网搜索引擎服务自律公约》，<https://www.isc.org.cn/hyzt/hyzt/listinfo-25501.html>，最后访问时间：2021 年 5 月 10 日。

21 例如，《反不正当竞争法》第十二条第二款：“经营者不得利用技术手段，通过影响用户选择或者其他方式，实施列妨碍、破坏其他经营者合法提供的网络产品或者服务正常运行的行为：…（四）其他妨碍、破坏其他经营者合法提供的网络产品或者服务正常运行的行为。”

22 参见苏青：《网络爬虫的演变及其合法性限定》，载《比较法研究》2021 年第 3 期。

便车”的行为便被认为是截取了平台经营者的竞争优势，损害其商业资源、妨害其正常服务，因而构成不正当竞争。²³

3. 涉嫌刑事犯罪

滥用“网络爬虫”技术，一旦达到一定的程度，就可能触犯刑法，就传统法益领域，可能侵害公民个人信息、知识产权等，就新型法益领域，网络数据系统安全等就可能成为“网络爬虫”技术侵害的法益。而判断以公民个人信息为对象的网络抓取行为是否入罪，关键要从整体上考量其对具体法益实质的侵害与威胁，以具体的法益保护导向来类化其入罪路径、界定其入罪标准。具体而言，在司法实践中，主要有两类刑事犯罪类型，即侵犯公民个人信息罪、侵犯著作权罪、侵犯商业秘密类犯罪及计算机类犯罪。其中，侵犯公民个人信息罪占比 24%，侵犯著作权罪占比 12%，非法获取计算机信息系统数据罪占比 12%，犯传播淫秽物品牟利罪占比 24%，其他类犯罪占比 28%，见图 7。

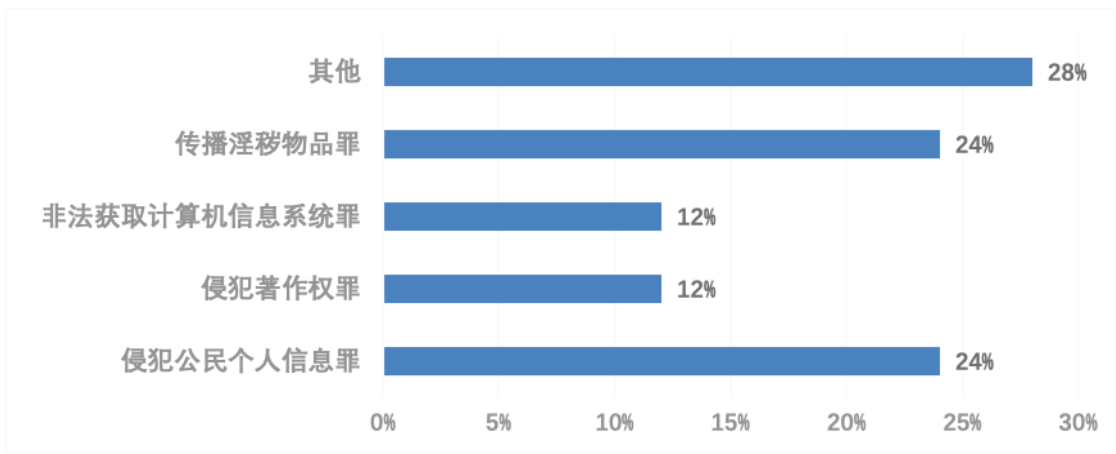


图 7：各种犯罪所占比例

(1) 侵犯公民个人信息罪、侵犯著作权罪、侵犯商业秘密罪。对于公民个人信息的侵害，从上文分析可以看出，如果只是针对个体，很难从民法层面对“网络爬虫”技术的使用进行规制，更多的情况是侵害的体量达到了一定的程度，直接入刑，以侵犯公民个人信息罪来予以认定。该条款是《刑法修正案（九）》新增的条款，具体规定为，“违反国家有关规定，向他人出售或者提供公民个人信息”，在司法实践中，也确实存在以该罪名定罪量刑的情形。

23 在谷米公司诉元光公司案中，元光公司利用“网络爬虫”技术获取了谷米公司开发的 APP“酷米客”中实时公交信息数据，并且将收集来的这些数据运用于本公司的 APP“车来了”中，向其用户提供了和“酷米客”APP 相同的公交信息的业务。法院认为，谷米公司通过收集公交实时数据，并进行分析、加工，提供给用户，可以帮助谷米公司在相关的市场占据优势地位，而元光公司的行为，坐享其成，削弱谷米公司 APP“酷米客”的市场优势，认定为元光公司构成不正当竞争。参见深圳市中级人民法院（2017）粤民初 822 号。

例如，魏江蒙案中，被告人魏江蒙利用“网络爬虫”技术，下载含有公民姓名和电话号码的工商个体户和单位的资料进行贩卖获利，情节特别严重，被判处有期徒刑四年。²⁴除了侵犯公民个人信息罪这一罪名，出现频率较高的还有侵犯著作权罪。如李金波等侵犯著作权案，被告人李金波注册水滴在线公司，后来设立网站“快读免费小说”，2014年起，被告人李金波、王强、徐文晖为了提高网站的知名度和点击量，在未获得玄霆公司许可的情况下，擅自通过“网络爬虫”技术复制、转载玄霆公司发行于起点中文网等网站上的文字作品，提供给其用户下载，获得收益。本案不是简单传统的“网络爬虫”技术的使用，而是定向的抓取数据，并将内置搜索引擎转码升级，使得盗版技术变异升级，案件更为复杂。²⁵

(2)非法入侵 破坏计算机信息系统罪。在黄后荣等非法获取计算机信息系统数据案中，福建微数公司软件工程师翁秀豪发现淘宝店漏洞，为增加派发优惠券的淘宝店数量，其向公司法定代表人黄后荣报告后，受到黄后荣的指示非法获取 cookie 数据，获取的淘宝用户 cookie 数据达到 2600 万组，利用获取淘宝用户的订单数据达到 1 亿条，情节特别严重，被定性为非法获取计算机信息系统数据罪。²⁶ 在新型的法益领域，“网络爬虫”技术的使用主要涉及的罪名有非法侵入、破坏、非法获取计算机信息系统数据罪等，具体侵害的法益是什么，需要通过对“网络爬虫”技术的具体使用过程以及使用后获得数据再进行利用的行为进行定性分析，方可得出结论。比如，运用爬虫技术加大服务器压力导致网站服务器崩溃可能构成破坏《刑法》第二百八十六条规定的计算机信息系统罪。²⁷

4. 行政处罚

司法实践中，针对违法利用爬虫技术爬取数据的行政处罚并不常见。据统计，截止目前，仅有 7 例相关的行政处罚案例。²⁸ 其中，基于爬取数据行为本身及后续的数据利用行为作出处罚的占比 28.6%，而大多数的案例是前述两种之外的违法行为模式，²⁹ 占比 71.4%，见图 8。

24 河南省济源市人民法院（2018）豫 9001 刑初字第 503 号。

25 上海市浦东新区人民法院（2014）浦刑（知）初字第 24 号。

26 杭州市余杭区人民法院（2014）杭余刑初字第 1231 号。

27 网络服务器是有承载限度的，频繁的大规模机器访问占用了原本服务器用于向普通用户返回数据的带宽和运算能力，如果不加控制地利用爬虫技术持续访问，实际上会影响到网络服务的正常运行，甚至使网站崩溃而无法访问，这也会影响到普通用户的正常访问。参见于娟、刘强：《主题网络爬虫研究综述》，载《计算机工程与科学》2015 年第 2 期。

28 见上述有关行政处罚案例的统计分析。

29 例如，某科技有限公司因不履行网络安全保护义务，而被行政处罚。参见京公房行罚字（2021）50913 号。

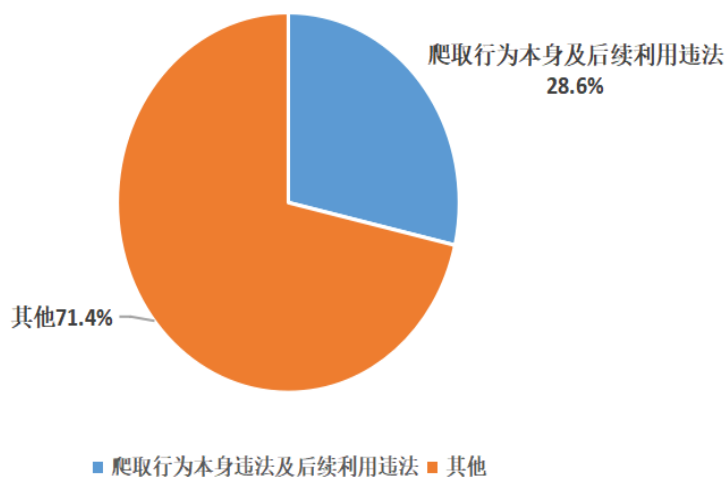


图 8：行政案件行为类型

司法实践中，公安机关、市场监督管理部门及证券监督管理机构等主要根据《网络安全法》《治安管理处罚法》及《计算机信息系统安全保护条例》等作出相关的行政行为。其中，根据《网络安全法》作出的占比 57.1%，根据《反不正当竞争法》作出的占比 28.6%，依据《计算机信息系统安全保护条例》作出的占比 14.3%，见图 9。³⁰

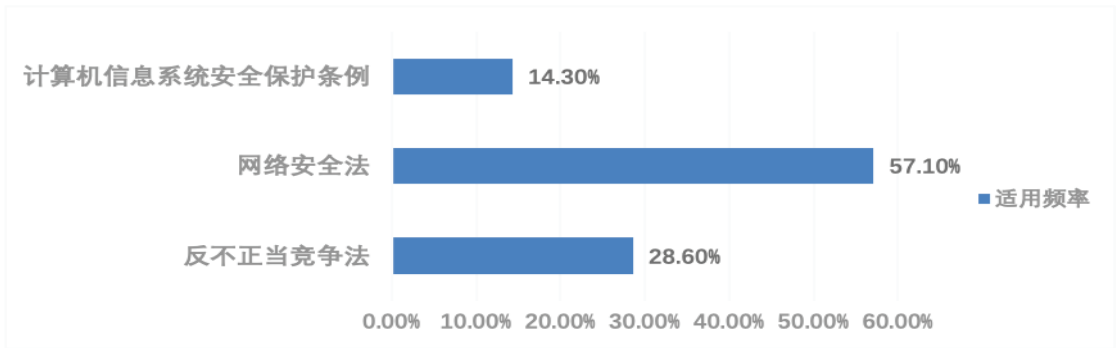


图 9：行政案件法律适用情况

例如，公安机关依据《网络安全法》《网络安全等级保护基本要求》中有关网络等级保护制度的要求作出的行政处罚。例如北京山顶洞商贸中心的网站存在代码执行漏洞，被北京市公安局房山分局认定为未落实网络安全保护义务，而依据《网络安全保护法》第二十一条及五十九条给予了警告处罚。³¹再如，中电标讯科技有限公司未采取防范计算机病

30 据上述统计分析，行政处罚方式主要是罚款和警告，分别占比 57.14%、42.86%。

31 京公房行政决字（2021）20915 号。

毒和网络攻击、网络侵入等危害网络安全行为的技术措施行为，亦被根据类似条款给予警告的处罚。³² 又如，农商帮网络技术有限公司因未依法落实计算机信息系统安全等级保护制度，致使其管理的网站被恶意程序控制，也受到类似的处罚。³³

（二）外部风险：基于爬取行为衍生的法律风险

根据前文对爬虫技术法律风险的行为类型化划分，下文将对爬虫技术使用者爬取数据信息后的后续使用行为所可能导致的法律风险进行分析，分析基于所获取的数据的性质以及数据本身拥有的权利保护规则，或者考虑数据作为竞争资源，使用数据的行为对市场竞争造成的影响。按照所爬取数据享有的权利保护属性的不同，数据使用行为法律风险分为以下三个方面。

1. 刑法规制范围内的数据使用行为

一般认为出于对特殊法益的保护，我国《刑法》对具有以下几种特定属性的数据内容予以规制。如果被传播的数据内容系淫秽物品，或用于牟利的数据内容系淫秽物品，或复制、上传的数据信息系著作权法保护的作品的，亦或是对公民个人信息实施的出售、非法提供行为，构成所规定行为后果的，则会存在构成传播淫秽物品罪、传播淫秽物品牟利罪、出售、非法提供公民个人信息等罪的刑事法律风险。在通过北大法宝检索到的 61 份与爬虫相关的刑事判决书中，以上述三类为案由的刑事判决文书总计为 39 份，占到了总判决文书数量的 64%，故此三类犯罪为爬虫衍生行为所面临的主要的法律风险。但实际上由于后续的行为方式多种多样，以致对其可能构成的犯罪无法周延列举，因此，还是需要根据具体的行为模式在具体的语境下进行判断。³⁴ 在排除爬虫技术本身使用行为所可能构成非法获取计算机信息系统数据、非法控制计算机系统罪以及由于属于特定属性数据内容的侵犯著作权罪、侵犯公民个人信息罪和复制、传播淫秽物品牟利罪外，基于爬取行为的衍生行为还可能构成诈骗罪、敲诈勒索罪乃至开设赌场罪，这三类犯罪的判决书总量为 15 份占到总数的 24%，其数量亦不可忽视。³⁵ 综上所述，基于爬取行为的衍生行为受到了我国《刑法》全面的规制。

32 京公房行罚决字〔2021〕50913 号。

33 京公西行罚决字〔2021〕50765 号。

34 参见陈军标、杨兰：“网络爬虫”技术的法律规制，载《东方法学》2021 年第 3 期。

35 见图 10。

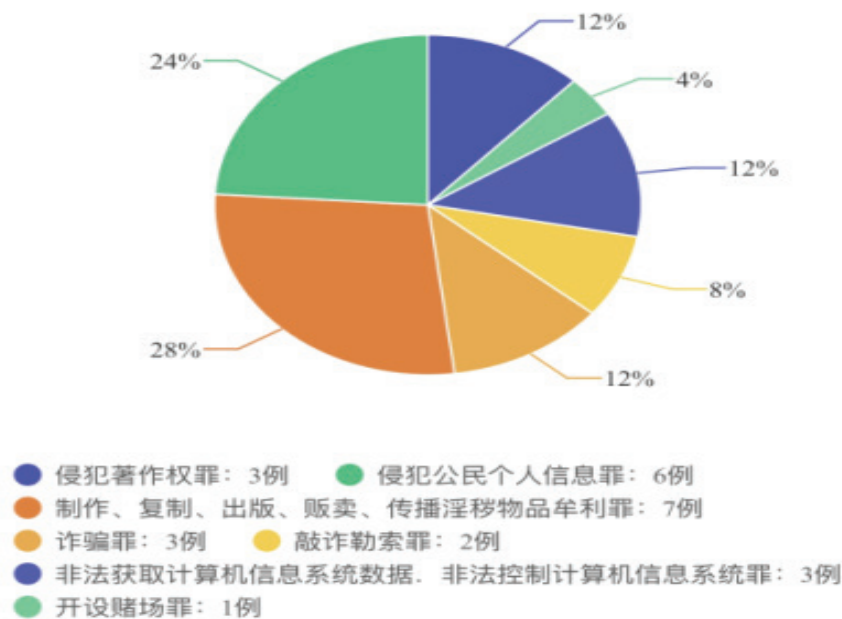


图 10：刑法规制下的爬虫行为各罪名占比

2. 知识产权法保护范围内的数据使用行为

在私法领域《民法典》和《个人信息保护法》均强调了对公民隐私和个人信息权保护，但个人信息的人格权属性使个人信息在民法上的实际保护存在一定障碍，往往发生的情况是对个人信息的侵害达到一定数量或者范围就直接入刑，而使其在民事领域的保护流于形式。故目前私法对爬虫的规制主要体现在知识产权法之上，而从现有案例来看，侵害信息网络传播权是最主要的法律风险，在北大法宝检索到的 290 份民事判决书中，以侵犯作品网络信息传播权为案由的判决高达 161 份。³⁶

毫无疑问，网络信息传播权是著作权中财产权的重要内容，其重点在于保护交互式传播方式下的著作权。根据《著作权法》和《信息网络传播权保护条例》的规定，侵害信息网络传播权的行为主要有两类行为：一类是内容提供行为，即信息网络传播行为，是指在向公众开放的网络中提供各类信息的行为。提供信息内容是最直接、最基础的传播行为，使得网络用户可以在选定的时间及地点获取作品。内容提供行为除在合理使用的情况下，不存在免责条款，即只要实施了传播行为即须承担侵权责任。³⁷ 另一类是网络服务提供行为，即为信息在网络上的传播提供技术、设备支持和中介服务的行为，包括自动接入、自动存储、自动传输、提供信息存储空间等。对于这类具有纯技术性的网络服务提供行为，《信息网络

³⁶ 见图 11。

³⁷ 参见芮松艳：《信息网络传播行为与网络服务提供行为的区分及侵权认定》，载《法律适用》2009 年第 4 期。

传播权保护条例》设立了避风港规则，网络服务提供者一般只有在明知或应知的情况下才承担责任。现实中，对于那些明显属于以上两类行为之一的判断并不困难。但是，具体到爬虫技术领域，其是网络的搜索链接服务，在实践中也有可能超越单独的技术服务提供行为，同样具有传播效果，对此需要作进一步区分。具体来说，需要区分提供的是通常意义上的搜索链接服务，还是将爬取内容存于自己的服务器中，进而提供涉案作品的在线下载服务。³⁸如果当事人仅提供一般的链接服务，并不提供具体内容，点击相应链接后显示的是被链网站，界面中展示的内容为被链网站中的内容。用户浏览的仍然是被链网站的内容，则不会受到侵权追责。如果利用爬虫技术以相对稳定和长期的方式将目标网页中的内容抓取并存储于自己的服务器中，在用户进行搜索操作时，从自己服务器中调取并提供给用户，用户在浏览器中浏览的是事先存储于自己服务器中的网页内容。³⁹即使提供的作品在播放过程中有页面播放源、视频水印图标等表象，但是，这种行为本质上仍属于《著作权法》意义上的复制，而非对原网页的搜索或者链接服务⁴⁰，因而其需要承担侵权责任。在中文在线数字出版集团有限公司诉深圳聚领威锋科技股份有限公司案中，深圳聚领威锋公司在其应用程序“石头阅读特别版”APP上显示：“实时追踪网络小说更新，并第一时间提醒，我们使用最新的网络爬虫技术，智能帮你追踪你喜欢的网络小说，作者更新后，笔者及时搜索整个中文网络，并整理好格式第一时间呈现给你……”，被告未经许可在该程序中向公众提供了涉案作品的下载服务，侵害了原告享有的信息网络传播权及获得报酬的权利，被法院认定为侵害原告作品网络信息传播权，必须停止侵害、赔偿损失。

38 北京卓易讯畅科技有限公司与央视国际网络有限公司著作权权属、侵权纠纷二审民事判决书，北京知识产权法院民事判决书（2015）京知民终字第2446号。

39 服务器标准是信息网络传播行为认定的合理标准。本院所持观点是：信息网络传播行为是信息网络传播权所控制的行为，对该行为的认定属于事实认定范畴，服务器标准最为符合信息网络传播行为这一客观事实属性。依据服务器标准，信息网络传播行为是指将作品置于向公众开放的服务器中的行为。需要特别指明的是，此处的“服务器”系广义概念，泛指一切可存储信息的硬件介质，既包括通常意义上的网站服务器，亦包括个人电脑、手机等现有以及将来可能出现的任何存储介质。参见北京易联伟达科技有限公司与深圳市腾讯计算机系统有限公司侵害作品信息网络传播权纠纷二审民事判决书。

40 北京三面向版权代理有限公司等上诉北京铁血科技有限责任公司著作权权属、侵权纠纷二审民事判决书，北京市第二中级人民法院民事判决书（2014）二中民终字第00919号。

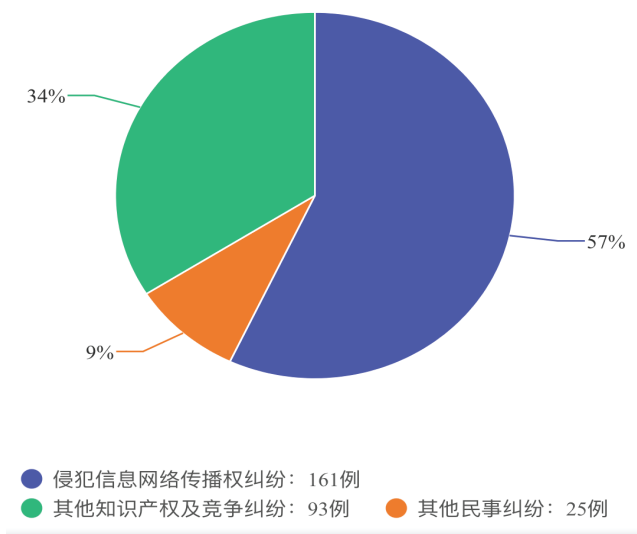


图 11：私法规制下的爬虫行为民事纠纷构成

3. 无明确权利属性的数据使用行为

有些数据既不属于公法保护范围，也无法落入到知识产权等私权的合理保护范围，但是，当这些数据是企业经营的重要投入要素，能够为企业带来市场竞争力。此类行为亦是法律规制的重点爬取竞争对手数据资源并使用的行为，其实质是竞争经营者之间的数据竞争行为。对于竞争对手数据资源的抓取和使用，原告一般主张被告行为构成不正当竞争行为。此时，司法实践中就需要从影响市场平竞争的角度，判断数据使用行为的法律责任。《反不正当竞争法》第九条规定经营者不得实施以盗窃、贿赂、欺诈、胁迫、电子侵入或者其他不正当手段获取权利人的商业秘密以及《国家工商总局关于禁止侵犯商业秘密行为的若干规定》都将侵犯权利人商业秘密的行为归类为不正当竞争行为。此外，除了属于商业秘密的数据使用行为，现实爬虫相关案例中构成不正当竞争的情形还包括，运用技术破解其他经营者服务器端获取数据并使用的行为、超越协议越权爬取竞争对手服务器数据的使用行为、通过爬虫软件截取其他经营者的用户流量导致广告收入受损等行为。如在（2017）粤03民初822号，原告谷米公司与被告元光公司、邵凌霜、陈昂、刘江红、刘坤朋、张翔不正当竞争纠纷一案中，法院认为被告元光公司未经许可，利用网络爬虫技术大量获取并且无偿使用原告谷米公司“酷米客”软件的实时公交信息数据的行为，实为一种“不劳而获”“食人而肥”的行为，具有非法占用他人无形财产权益的不正当目的，谋取该软件在实时公交信息查询软件中的竞争优势，谷米公司的上述行为违反了诚实信用原则和公认的商业道德，构成不正当竞争。

三、“对象类型化”下“网络爬虫”的合规路径

通过上述分析，网络爬虫技术的应用会产生刑事、民事、行政等多方面的法律风险。基于此，对“网络爬虫”的使用进行全面合规是信息化社会的应有之义。其原因在于，无论是数据权属与不正当竞争的场景性界定，还是大数据与隐私保护的矛盾与张力，网络爬虫抓取数据的正当性、有效性都与数据的访问权限和开放程度密切相关，一旦突破一定的边界就会诱发风险。⁴¹从立法者对信息和数据从人格权和财产权进行分置的安排可见，信息与数据存在区别。《个人信息保护法》第四条第一款与《数据安全法》第三条第一款分别关于个人信息和数据的定义，则进一步指明了二者的区别所在。⁴²信息是内容、知识等，其功用在于解决不确定性；而数据是形式，是表现信息的载体。⁴³在“网络爬虫”的法律规制情境中，信息与数据皆可能成为被爬取的对象，而数据因其记录和存储信息的属性往往成为海量信息被爬取的介质。本文将爬取对象作类型化划分，基于前文数据分析对个人信息、政府数据和企业数据三方面“精准施策”，提出以下合规建议。

（一）爬取公民信息、隐私、作品的合规建议

1. 公民个人信息

在统计的民事案件样本中，利用网络爬虫技术爬取公民个人信息的案件占比达 43.33%，其中还不包括既爬取企业信息又爬取个人信息的案件数量。在行为类型的划分下，爬取行为本身违法的仅占 6.67%，爬取数据后续利用行为违法的则高达 90%。对于公民个人信息来说，网络爬虫抓取公民个人信息行为是否合法的根本在于是否获得了信息主体的真正授权同意，以及在同一的范围内合理实施。刑事案件样本亦然，爬取公民个人信息的犯罪比例占 28%，在行为类型划分下，爬取行为本身违法的占 16%，后续使用数据违法的占 24%，两者兼具的占 52%。由此可见，针对爬取公民个人信息的行为，应当尤其注意是否获得授权同意，以及后续数据的合法使用。有鉴于此，可从以下角度构建合规规则：一方面，在避免过度干涉数据共享与数据流通的前提下，应当避免在未取得用户个人授权同意的情况下爬取，即使是爬取网站公开的个人信息，也需要审查该数据的授权情况，包括个人信息来源的合法性、授权同意转让、共享情况等。不能仅以该数据系公开信息而未经用户同意直接爬取。同时，根据《网络安全法》《个人信息保护法》《个人信息安全规范》，爬取个人信息应遵守合法、正当、必要之原则，避免因过度爬取干扰信息系统的正常运行或提高个人信息泄露的风险。在爬取信息后，如在后续业务开展中所将进行的个人信息处理活动已

41 孙禹：《论网络爬虫的刑事合规》，载《法学杂志》2022 年第 1 期。

42 参见申卫星：《数字权利体系再造：迈向隐私、信息与数据的差序格局》，载《政法论坛》2022 年第 3 期。

43 参见申卫星：《论个人信息权的构建及其体系化》，载《比较法研究》2021 年第 5 期。

超过前序授权范围，则应当重新获得信息主体的同意。另一方面，数据采集者接到信息权利人关于停止采集个人信息通知的，应立即停止采集行为，并应要求采取删除已采集数据等补救行为。

2. 公民隐私

隐私与个人信息的区分在于隐私权只具有防御作用的消极权能，而个人信息具有传播、利用等积极权能，⁴⁴ 因此对于爬虫技术的规制手段和程度应作区分。针对爬取公民隐私的行为，可从以下几个方面予以规制。首先应禁止在未经授权同意之情形下爬取他人隐私甚至是爬取后加以利用。其次，应当提高医院、学校等掌握海量公民隐私的单位的信息系统的“准爬”和“反爬”的技术防范标准，如果爬虫行为因过高频次的访问或爬取妨碍了网站或计算机系统的正常运行，甚至导致系统存储的涉及公民隐私的信息处于暴露的高度风险，例如对医疗机构的网络系统进行爬取，导致其系统崩坏而使得患者的就诊信息处于受到黑客攻击盗取的高度风险中，则会造成难以预估的恶劣影响。最后，值得注意的是，对于大量掌握此类信息的主体而言，其应负更高的注意义务，在该部分的合规建设和内部管理应投入更多，在做好对外防范工作的同时应注重法人内部治理，增强内部人员的合规意识。

3. 作品

利用网络爬虫技术爬取作品，在民事领域具有侵害信息网络传播权的较大法律风险，在北大法宝检索到的 290 份民事判决书中，以侵犯作品网络信息传播权为案由的判决占比 56%，该权利是著作权中财产权的重要内容。在刑事领域则很可能构成侵犯著作权罪，其中侵犯著作权罪在全部刑事爬虫案件中占比达到 12%。所以，讨论爬取作品时的风险应对措施具有较大的必要性。对于作品来说，如果网络爬虫在爬取公开数据时避开或者破坏了保护著作权的技术措施，则很可能因违反前置性保护要求而构成侵犯著作权罪。如李金波等侵犯著作权案，被告人李金波注册水滴在线公司，后来设立网站“快读免费小说”、2014 年起，被告人李金波、王强、徐文晖为了提高网站的知名度和点击量，在未获得玄霆公司许可的情况下，擅自通过“网络爬虫”技术复制、转载玄霆公司发行于起点中文网等网站上的文字作品，提供给用户下载，获得收益。因此，网络爬虫构成侵犯著作权罪的风险主要在于爬取数据时是否突破了限制浏览作品的技术措施。有鉴于此，针对网络爬虫爬取作品涉及侵犯著作权罪的风险，合规规则的构建应当注意，当网站管理者基于著作权保护的目针对相关内容设置技术措施时，不应突破技术措施的限制爬取相关内容。⁴⁵

44 参见谢远扬：《个人信息的私法保护》，中国法制出版社 2016 年版，第 24 页。

45 参见孙禹：《论网络爬虫的刑事合规》，载《法学杂志》2022 年第 1 期。

（二）爬取政府数据的法律风险应对

《数据安全法》对国家机关公开政务信息数据做出了明确要求。爬取方侵害政府的计算机信息系统安全、重要数据安全或者国家秘密中必然是被禁止的。若爬取的是政府公开的信息，因暴力爬取行为导致被访问的服务器负荷过高，当大量网络爬虫同时访问网站则会大量占用宽带流量，最终造成服务器不堪重负直至崩溃，从而使目标计算机或网站无法提供正常服务，⁴⁶爬取者也可能因此成立破坏计算机信息系统罪。针对网络爬虫爬取政府数据的风险，应当对可能涉及侵犯国家秘密罪在选择爬取对象时应当保持高度警惕，注意甄别爬取对象，禁止对涉密国家事务、国防建设、尖端科学技术领域的网站和计算机信息系统进行爬取。而对于暴力爬取数据并破坏信息系统的行为，一方面，爬取方应当根据网站的承受能力、访问情况来限制数据爬取的速率，或者将爬取速率控制在最低范围，例如每秒10-15次；还可以采取控制数据爬取的时间段的方法来规避，在网站访问的非高峰期实施爬取数据，此外，还可以采取必要措施以避免对服务器产生影响，如编写网络爬虫时提供可识别的信息，以便网络运营者因数据爬取遭受影响时联系数据爬取方，要求其调整爬取的速率、时间或者停止爬取行为。

面对市场主体对信息要素日益扩张的需求，也出于降低相应的针对网络爬虫违法行为的执法成本，政府机关应致力于提升公共数据供给及智能公共服务能力，进一步落实国务院《政务信息系统整合共享实施方案》等，综合运用信息手段建成支撑各类政务服务应用程序的精准高效、动态实时、集中统一的行为数据资源库。⁴⁷

（三）爬取企业数据的法律风险应对

1. 作为供给端的法律风险应对

上文分析表明，对爬取的数据进一步加工利用，转化为经济效益才是产生绝大多数爬虫技术行为侵权的目的。在对爬取申请进行授权，向他人提供数据时，应对接收方进行充分的尽职调查，在合同管理的免责声明和廉洁承诺等方面强化管控，防止产生帮助网络信息犯罪的刑事风险。涉及跨境业务的企业则应当留意尚未生效的《网络数据安全条例》关于数据跨境传输的相关规定，把握其制订动向，从而对自身的数据合规建设作出调整。

2. 作为需求端的法律风险应对

（1）爬虫技术行为本身的使用风险应对。一是准入环节。如上文分析，无视 robots 协议而随意抓取网站内容行为和绕过或破坏反爬措施爬取数据的行为二者合计占爬虫不正当竞争案例逾七成。面对这两类“高发”事由，企业严格遵循 robots 协议或者与权利方的合

46 参见张一献：《从技术到犯罪：恶意网络爬虫行为入罪的类型认定与裁判思路探索》，载《时代法学》2020年第4期。

47 参见马长山主编：《数字法治概论》，法律出版社2022年版，第125页。

同约定，一般不构成民事违法。若为单方授权型协议，则需明确权利告知事项和权利的行使边界；若为合同，则须注重合同的内容本身有效性及相关条款的严格遵守、履行的可行性。另外，应避免（故意）突破，甚至是通过伪造身份信息、破解加密规则等破坏网站的防护措施，或非法获取权限进行爬取，这方面尤其需要做好法人内部管理，防止企业内部人员忽视合规要求或将业务擅自外包给第三方机构完成产生的法律风险。

二是行为限度。合理控制爬取行为及限度，以免造成爬取对象网站过载甚至停止运行。参考暂未生效的《数据安全管理办法（征求意见稿）》，建议爬取数据流量不超过网站日均流量 1/3，或如前文爬取政府数据的法律风险应对所提及，可以“错峰”爬取。总之，在这一方面应通过加大技术投入做到及时、准确地监测爬取行为的限度与边界，并通过优化自身的代码，避免干扰被访问网站的正常运行。

（2）爬虫技术行为对象的合规风险应对——以商业数据为例。一是竞争关系的认定及风险评估。对于商业数据的爬取应注意是否存在“搭便车”“不劳而获”的不正当竞争行为，评估自身业务与被爬网站是否存在竞争关系，爬取行为是否存在分流用户、提供替代性服务，并导致被爬取对象用户吸引力降低、竞争优势削弱的后果，如是，则需慎重使用爬虫技术。具体而言，对于需要抓取同类型或与自身有竞争关系企业的网站信息，应当注意考量对爬虫信息的使用范围和方式以及对被爬企业造成的影响，考量对数据的使用是否会对被爬企业的业务造成实质性替代，是否会造成其市场份额和客户的减少、造成被爬企业现实的或预期的利益损失，并谨慎利用所爬取的信息。

二是商业秘密侵权的风险防范。有统计显示，当前科技公司约 60% 的创新成果最初都是以技术秘密的方式存在的。一旦以海量数据为载体的商业技术秘密遭到爬虫技术的非法窃取，造成损失的速度和体量都是难以预测的。以我国司法史上判赔额最高的侵害商业秘密案件——“香兰素”技术秘密侵权案为例，判赔数额奇高的背后是近 8 年的长期处于侵权状态下企业自身合规监管缺失和双方历经 3 年的诉讼纷争。因此，企业应对照国家市场监督管理总局《商业秘密保护规定》对商业秘密内涵和外延的规定，以及侵犯商业秘密行为的构成，完善自身的技术资料获取方式的监管制度，在每一次具体完成抓取行为后应做到积极审查，一旦发现侵权信息应及时停止抓取并彻底删除，防止因时间的延误造成损失的不必要扩大。

总之，大数据被认为是引领未来发展的重要技术变革，网络爬虫作为数据收集、流通的重要方式，则是大数据时代不可或缺的一个角色。但是，爬虫技术应当在一定的制约下使用，否则即便是中立的技术手段也会给使用者带来法律风险甚至是刑事责任。

（四）全面合规：裁判规则的实证总结

通过提炼法院对爬虫案件的裁判规则，具体包含民事侵权、不正当竞争和刑事犯罪类案件，尝试建立起一套全面的“网络爬虫”的合规体系。⁴⁸ 其中，在民事侵权类案件中，“遵照合法、正当、必要的爬取原则”占比最大，达到了 55.4%，见下表 1。

民事侵权：主要是侵犯公民个人信息权益		
	合规路径	占比
（一）	爬取原则：合法、正当、必要	55.4%
（二）	从知情同意看，“授权”不等于“完全授权”，不得超出“授权”的界限，否则不属于知情同意	33.7%
（三）	公民的敏感个人信息必须取得单独同意	24.3%
（四）	公民的私密信息属于隐私范畴，不得收集和处理	17.5%

表 1：民事侵权的合规路径

在不正当竞争类案件中，“遵守商业道德和行业规范”显得尤其重要，其占比达到了 55.7%；其次便是“不劳而获、搭便车，实质性替代竞争对手的市场，造成被爬企业现实的或预期的利益损失”，占比 47.6%，见下表 2。

不正当竞争		
	合规路径	占比
（一）	禁止爬取商业秘密（判断标准：秘密性、价值性、保密性）	25.2%
（二）	遵守 robots 协议，禁止破坏反爬措施	27.5%
（三）	遵守商业道德和行业规范	55.7%
（四）	不得不劳而获、搭便车，实质性替代竞争对手的市场，造成被爬企业现实的或预期的利益损失	47.6%

表 2：不正当竞争的合规路径

而在刑事犯罪类案件中，对“侵犯著作权罪”、“侵犯商业秘密罪”及“侵犯公民个人信息罪”等六个主要罪名进行了区分。比如，在“侵犯著作权罪”中，“未经权利人许可、

48 “占比”意为在同类型案件中，某一法院采用该裁判规则的情况占该类案件总量的比例。

授权而通过复制、下载作品等，足以妨害所有权人实现排他性处分、使用收益权能”为主要入罪标准，占比达到了 85.1%，见下表 3。

刑事犯罪			
基本标准	客观标准：有突破数据保护措施的行为	具体标准：考察被抓取的数据类型及网络爬虫抓取数据行为对法益的侵害或者威胁是否达到实质可罚的程度	
	主观标准：行为人有突破数据保护措施的故意	具体标准：认识到突破数据网站技术措施的行为违背了权利人的保护意愿，仍选择继续爬取数据	
	罪名	入罪标准	占比
(一)	侵犯著作权罪	未经权利人许可、授权而通过爬虫技术复制、下载作品等，足以妨害所有权人实现排他性处分、使用收益权能	85.1%
(二)	侵犯商业秘密罪	未经授权突破保密措施，爬取以数据为载体的商业技术秘密，权利人因此承受销售利润大幅下降的损失	79.5%
(三)	侵犯公民个人信息罪	涉案个人信息未彻底“去标识化”处理，行为人仍存在实现“重标识”的技术可能性	67.4%
(四)	非法侵入计算机信息系统罪	未经授权进入涉及国家安全和国家秘密的政府内网、国防建设、尖端科学技术等特殊领域且依法无授权外部访问的计算机信息系统	89.8%
(五)	非法获取计算机信息系统数据罪	侵入前款规定以外的计算机信息系统或者采用其他技术手段，获取该计算机信息系统中存储、处理或者传输的数据的行为	92.1%
(六)	破坏计算机信息系统罪	以爬虫技术为工具，逾越一般计算机信息系统可承受的爬取限度，对计算机信息系统或数据进行破坏，或对计算机信息系统安全措施进行暴力破解	82.9%

表 3：刑事犯罪的合规路径

四、结语

以司法裁判文书为研究对象进行实证分析，可以发现：“网络爬虫”引发法律风险涉及民事侵权、不正当竞争及刑事犯罪等多个维度；从“网络爬虫”与法律风险的关系看，不仅“爬取”行为本身具有高风险性，而且爬取后续的使用行为更是使风险蔓延地更加严重。此外，学界对爬虫行为规制的立法论研究中，大多从数据确权的角度进行分析，通过赋予被爬取主体排他性的权利来防止爬虫行为。但是，爬虫绝不仅仅是“害虫”。在数据要素市场建立的过程中，不能一味阻止数据的流通。因此，与其在立法上诉求数据确权，不如从行为规制层面入手。具体而言，应当立足于现有规范，爬取主体应根据爬取的对象（公民、政府、企业）明晰主体义务，在履行义务的基础上、在合法的范围内进行爬取。未来，随着爬虫行为的规范化，数据的市场潜能必将得到释放，这将有助于后疫情时代经济的复苏。

Legal Risks and Compliance Paths of “Cybercrawler”: An Empirical Study Based on 242 Judicial Judgment Documents

Yong Zhang Qianyi Yan

Abstract: While “Nice Cybercrawler” promote the innovation of business models and the establishment of data market, “Evil Cybercrawler” cause security risks for citizens, enterprises and government information data. Nowadays, the criminal regulation of online data crawling behavior has become the fact in the judicial practice and is gradually expanding. However, previous research on the legal regulation of “Cybercrawler” has mostly focused on theoretical analysis, lacking empirical verification, and failing to draw its governance spectrum using a “panoramic” observation paradigm. In this context, by conducting statistical analysis of 242 judicial judgment documents and using empirical analysis research methods, the legal risks of criminal crimes, administrative violations, and civil torts related to the “Cybercrawler” behavior itself and subsequent behaviors were separately sorted out. And based on the type of data crawled, compliance paths have been developed for the “Cybercrawler” behavior of individuals, enterprises, and government data, hoping to clarify the compliance path for the global governance of “Cybercrawler” and provide useful reference for relevant experts, scholars, and colleagues in the practical field.

Keywords: Network Crawler , Criminal Offenses , Civil Infringement , Administrative Violations , Compliance Path