文章编号：2095-3747（2023）-04-0038-05

大数据时代下网络爬虫的刑事入罪规制

李佳晓

（上海政法学院刑事司法学院，上海 201710）

摘 要：大数据时代下，数据信息呈爆炸式增长，网络爬虫作为一款数据搜索引擎应运而生，但在其使用过程中相关数据安全法益遭受到威胁。纵观实践中有关爬虫行为涉及的案件，以相关计算机信息系统被侵害案件与相关公民个人信息被侵害案件居多；而前者所涉及的罪名中，涉嫌非法侵入计算机信息系统和非法获取计算机信息系统数据的罪名较多。网络爬虫作为一项中立性的技术工具，只有在被划归为恶意爬虫行为时才需由刑法予以规制，而由于法律本身存在滞后性的缺点，恶意爬虫行为的入罪路径需要进一步明晰。总体来说，对于恶意爬虫之行为入罪不能脱离基本的定罪原则即主客观相统一原则和罪刑法定原则。

关键词：网络爬虫；恶意爬虫行为；主客观统一原则；罪刑法定原则

中图分类号：D923　　**文献标识码：**A

新世纪以来，人们通常借助于互联网来储存个人信息。同样，公司、企业和其他社会组织的大量信息也依托于互联网进行存储、传播、共享。在杂乱无章、繁冗复杂的数据信息中，如何汇集以及如何搜索信息成为了一个难题。为此，网络爬虫应时而生。网络爬虫，一般叫网络机器人，也叫做网络蜘蛛 (InternetSpider)，因为其工作运行模式与“蜘蛛”相类似。网络爬虫本质上作为一种程序，可以根据设定好的规则，按照事先规制的路径，模拟人工点击的操作方式在以互联网为媒介的各种小程序、APP、网页中“扒取”数据，并且将获得的数据信息进行分类和再提取，以便于更精准地向使用者投放。作为互联网时代下衍生的工具，这项技术被广泛使用，许多互联网公司通过获取用户的位置、喜好等个人信息，从而精准地狙击用户的喜好，在此基础上引导消费者购买产品，达到产品推广的效果，从而获利。但部分不法分子从中看到“商机”，通过侵入互联网公司系统的方式而获取数据，损害他人利益。如深圳市某米公司（以下简称“某米公司”）与武汉市某光公司（以下简称“某光公司”）在软件经营开发中存在竞合。某米公司是“酷米客公交”APP 的开发者以及运营者，某光公司是“车来了”APP 的开发者和实际的运营者。两款软件的主要功能都是为用户提供地理定位、公交线路查询、线路规划、实时公交地理位置等服务，免费供用户下载、使用。某米公司通过安装 GPS 相关设备可以获得公交车的实时信息并通过后台予以发送，用户利用 APP 查询时可以立马收到信息反馈。公交作为公共交通工具，其实时运行路线、运行时间等信息只是客观事实，但当这类信息通过人工采集、分析、编辑、整合并配合 GPS 精确定位后，APP 提供的数据越准确，用户收到的信息反馈越及时，意味着该软件将占有更大的市场份额。某米公司凭借此款 app 以其高度的精准性在同类竞争市场中获得了较高的市场地位与优势。某光公司意识到了这一点，为了提高己方 App 的市场占有率，便利用网络爬虫软件窃取某米公司 App 相关数据，并投射于己方 App 中供用户使用，以此获利，[参见广东省深圳市南山区人民法院〔2017〕粤 0305 刑初 153 号刑事判决书。] 从而对某米公司的合法权益造成了侵害。

面对层出不穷的“网络爬虫”案件，为了维护正常社会秩序，保障国家和人民安全利益，我

收稿日期：2023-02-20
作者简介：李佳晓（2000—），女，硕士研究生

国民法、行政法等都出台了相关规定对以使用“网络爬虫”进行违法犯罪的行为予以规制。

但随着技术性的不断提高以及互联网快速发展带来的巨大流量与数据红利，某些爬虫行为的社会危害性已经触及到社会法益的底线，需刑法予以规制。

在刑法意义上，学界一般以网络爬虫是否违反爬虫协议（通常指数据合法的持有者为了保障数据不被第三方非法获取，通常会以明示或是默示的方式与第三方签订一个契约，契约的目的在于保护数据的秘密性。）为依据，将网络爬虫分为善意的爬虫与恶意的爬虫两类。善意的网络爬虫是指，不违反爬虫协议的有关规定，遵循一定的规则，在获取网站数据的前提下，不影响网站的正常运营，数据的获取范围和获得方式也应当遵循爬虫协议。恶意的网络爬虫是指，违反了“robot”协议事先约定好的规则，行为人恶意使用爬虫程序进入网站获取数据的行为。[1] 需刑法予以规制的自然应当是恶意的爬虫。

1 网络爬虫罪与非罪的界限划分

上述某光公司使用爬虫软件爬取某米公司计算机系统数据的行为就是爬虫行为。类似行为还有上海某品网络技术有限公司获取数据行为。北京字节跳动网络技术有限公司、上海某品网络科技有限公司都经营技术开发、技术服务、电子商务电子制造品等计算机网络技术领域的业务。某品网络科技有限公司高层张某某、宋某某、侯某某采用技术手段抓取被害单位北京字节跳动网络技术有限公司服务器存储的视频数据，并由侯某某指使郭某某破解北京字节跳动网络技术有限公司的防抓取措施，使用“tt_spider”文件实施视频数据抓取行为，致使被害单位北京字节跳动网络技术有限公司损失技术服务费人民币 2 万余元。[参见北京市海淀区人民法院〔2017〕京 0108 刑初 2384 号刑事判决书。] 上述行为是否构成犯罪？刑法是否需要对上述爬虫行为予以规制？如何规制？规制的标准是什么？学界对此展开了激烈的讨论。

从朴素刑法观的角度出发，人们评价新生事物是否应当为刑法所保护或规制的基本立场在于这一事物是否具有值得为刑法所保护的价值，以及是否具有对公共安全或他人安全造成侵害或威胁的可能。现代刑法学理论中又将朴素刑法观中认定刑法是否应当保护或规制的内容抽象为法益一词，以是否侵犯刑法所保护的法益作为评价标准。针对网络爬虫而言，其是否应当受刑法规制，在于其是否侵犯了刑法所保护的法益。[2] 对于上述某光公司案和某品公司案，刑法对其作出的评价就是两种行为都损害了刑法所保护的法益，即我国的计算机信息系统安全。

从风险刑法观来看，我国目前已进入风险社会。在风险社会中，刑法立法应以控制风险的需要为原则，以维护刑法规范的有效性为出发点，寻求自由与安全的平衡。这是对风险刑法观的基本要求，也确立了社会本位的和谐。风险刑法观并不以提倡刑法扩张为己任，而是直面刑法扩张的事实，说明其社会背景和原因，基于风险提出兼具现实性和建构性特征的“社会动荡”和“不安”等结构的风险而不使用刑法，试图确定刑法扩张的合理限度。[3] 爬虫技术作为新兴技术，在应用过程中必然面对不可预知的风险，尤其是行为人在主观恶意的情况下违背数据管理者的意愿强行爬取数据，必然会对网站、计算机系统造成不同程度的破坏风险。倘若数据管理者设置了反爬虫措施，则行为人暴力爬取数据造成的风险程度更甚且更不可预测，由此，刑法必然要积极地预防，保持社会的稳定与和谐。

从功利主义的刑法观来看，一方面，对有必要的恶的犯罪行为进行的处罚给犯罪人带来的不利影响应该大于该犯罪所得；另一方面，有利于增进社会的幸福总量，必须充分保障公民对个人幸福的追求。[4] 从第一个方面来看，由于恶意爬虫行为对现实社会和个人造成了利益侵害，例如上述某光公司案和某品公司案，此是对社会和个人之恶，如果只依赖于民法或经济法，所达成的社会效果不足以抵制此种行为造成的恶果，因此应当由刑法加以规制才能使刑罚收益大于其对犯罪人造成的伤害。对于第二个方面，有刑法的严厉性和最后手段性作为保障，一是可以增加一般群众对社会信赖之利益；二是可以为技术开发者提供足够的空间进行技术创新，促进技术发展；三是对潜在的危险者起到一定的警告作用以消除他们可能使用爬虫技术而给社会带来的危害性。

2 网络爬虫定罪原则

2.1 主客观相统一原则

爬虫行为有罪判决要遵循主客观相统一的原则。这实际上是以我国刑法犯罪构成中主观要件和客观要件的有机统一为基本逻辑支撑点，是犯罪成立层面应遵循的原则。

第一，主观方面表现为行为人故意为之，即行为人明知自己的行为具有社会危害性而希望或者放任这种危害结果的发生。首先，关于“明知”应当有一个清楚的界定。在司法实践中，对“明知”是否包含违法性认识存在争议。由于爬虫技术的新颖性以及有关立法的滞后性使得行为人在使用网络爬虫时并不明确刑法规制的边界。朴素刑法观认为是否具备违法性认识并不影响对行为归罪。换言之，法律公布之后就应当推定为一般人已经知晓，以防止部分犯罪人以此为借口实施危害社会的犯罪行为。然而，爬虫技术本身的复杂性和新颖性由刑法定性对法律人来说存在困难；同时由于技术使用的简便性，如果刑法对行为入罪的边界不明晰很容易造成罪名适用过宽或过窄，既不利于实现刑法特殊预防的目的，也不利于社会经济的持续向前发展。[5] 就目前来说，对于行为人“明知”的认定应当着重考察行为人对危害结果的认识以及行为人的主观目的。如果行为人一开始就是抱着为了实现某些特定犯罪而爬取数据或者行为人清楚地知道或者以无所谓的态度暴力爬取网站数据都应当属于“明知”；而如果是为了科研或教学等合理使用的目的而进行爬取行为那么可以认定行为人的动机纯洁，当然仍需要考察行为人爬取行为是否给网站的实际运营带来了危害。

第二，客观方面表现为恶意的爬虫行为，判断标准主要在于对“突破反爬虫措施”的性质认定。“反爬虫措施”包括“robot 协议”以及反爬虫技术壁垒。违反了前者应当由民法等前置法予以规制，而突破反爬虫技术壁垒应当由刑法予以规制。

2.2 罪刑法定原则

罪刑法定原则的基本含义包括：认定行为人的行为构成犯罪和给予处罚，必须以刑法的明文规定为前提；如果刑法没有明文规定，即使行为危害很大，也不能认定犯罪和给予处罚。即法无明文规定不为罪、法无明文规定不处罚。对于恶意的爬虫行为是否应当入罪，入什么罪，应当以刑法总则为方向，以刑法分则条文为路径出发。通过上述分析可以确认恶意的爬虫行为应由刑法予以规制，但具体入什么罪应当进一步讨论。

（1）侵害计算机信息系统安全类案件

在有关爬虫行为侵害计算机信息系统安全类案件中，涉及的主要罪名包括非法侵入计算机信息系统罪和非法获取计算机信息系统数据罪。两者的争议点都聚焦在行为本身即何为“侵入”以及法益“计算机系统”与“数据”。刑法中“计算机信息系统”是指计算机、网络设备、通信设备、自动控制设备等具有数据自动处理功能的系统，这一点在实践中常常不难认定。所谓非法侵入，是指违反国家计算机信息系统安全保护法律、法规，未经国家有关部门合法授权或者批准，擅自通过计算机终端访问或截取国家重要领域计算机信息系统的数据。在司法实践中，由于对“侵入”和“数据”本身的界定较为模糊，从而导致非法获取计算机信息系统数据罪逐渐沦为计算机系统犯罪尤其是爬虫行为入罪规制的“口袋罪”。要落实罪刑法定原则，“去口袋化”，就必须对两者的概念做一个清晰的界定。首先，要明确“侵入”行为的本质在于非法访问计算机系统。但对非法访问含义存在争议。有学者认为，“非法访问”单指“未经授权”；有学者认为“未经授权”与“超越授权”都包含于此。从立法本身出发，《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》第 2 条分别对“侵入”计算机信息系统的程序、工具作出了界定：“具有避开或者突破计算机信息系统安全保护措施，未经授权或者超越授权获取计算机信息系统数据的功能”。

持“未经授权”一元论者认为，此项解释是我国在借鉴国外立法的基础上提出来的，因此，还原当时的立法背景，侵入犯罪立法所规定的行为方式为“未经授权”+“访问”；并且对于非法获取计算机信息系统数据罪来讲，侵入行为如果包含获取行为，那么在刑法体系与逻辑上势必造成混乱，而立法者后来再单独设立此罪名的意义就在于表明获取数据行为本身的危害性，与此前非法侵入计算机信息系统罪相区分。二元论者认为，“超越授权”主要有两种：一是纵向超越授权。它是通过一定的技术手段，获得更高的权限来控制计算机信息系统。二是横向超越授

权。即行为人虽然获得了一定的准入权限，但却在权限范围或时间之外。[孙杰．数据爬取的刑法规制 [J]．政法论丛，2021(03)：115-125.] 在一起检察院不予批准逮捕案中，于某作为公司人员，通过公司内部聊天软件利用爬虫技术获取公司内部员工个人信息。检察机关最终作出不逮捕决定的主要原因是：作为某公司的正式员工，其爬出的数据信息，本在其职责范围内是知情的对象，并没有超出自己的权限范围，也没有以“非法侵入”来排除刑事责任。另外，根据公司的规定，某人可以注册一个聊天工具，阅览和查阅公司的员工资料信息。由此可以证明“非法访问”应当包含“超越授权”。其次，“非法侵入”的具体判断标准是什么？主要观点有合约授权理论与“代码理论”。后者为主流观点，前者是从“授权”的字面意思出发，认为以网站单方面授权的意思表示以及网站和相对方的契约作为标准。这反映了数据网站对数据“微弱”的保护意愿，是披着契约外衣的“私规”，在技术上起不到强制作用。代码理论体现了数据主体对数据的“强烈”保护意愿，基于技术措施确定网络爬虫的行为边界，可以在用户隐私保护与数据共享之间划定清晰的界限。抓取者在明知没有授权的情况下，故意避让或强行突破技术措施，属于对数据的“未经授权”访问或获取。[杨志琼．数据时代网络爬虫的刑法规制 [J]．比较法研究，2020(04)：185-200.][6] 最后，有关数据和信息的划分需要进一步明晰。有学者认为，数据犯罪所侵犯的法益是数据安全，包括数据的保密性、完整性和可用性等方面。另有论者主张，数据安全法益的本位在于对信息内容的保护而不是对数据载体的保护，数据的保密性即数据所承载的信息内容的保密性。[7] 数据通常是信息的载体，其在计算机系统中通常表现为符号代码，而不具有实体性；相对地，信息通常是指具体事件的内容、某项理论等有形体，所以数据与信息本质上存在区别，不能混为一谈。在网络犯罪中，数据安全法益应当作为一项独立法益而为刑法所保护。由于数据与信息本质不同，所以两者的安全概念也存在区别。前者主要是指操作技术层面的安全性，而后者则侧重于具体内容是否被他人所知的安全性。网络爬虫作为一类技术性软件，其法益应当为数据安全而非信息安全。

（2）侵犯公民个人信息案件

通过明确恶意爬虫行为的行为类型，在使用爬虫技术侵害公民个人信息的入罪路径上，应当对“个人信息”作出清晰的界定。2016 年网络安全法首次以法律的形式正式给个人信息进行了定义。第七十六条规定：“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自然人个人身份的各种信息，包括但不限于自然人的姓名、出生日期、身份证件号码、个人生物识别信息、住址、电话号码等。”《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》第 1 条规定，“刑法第二百五十三条之一规定的‘公民个人信息’，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、身份证号码、通讯联系方式、住址、银行账号、财产状况、失踪轨迹等。”相对于仅将“可识别性”确立为公民个人信息判断标准的《中华人民共和国网络安全法》，《个人信息刑事解释》指出，“公民个人信息”在刑法意义上是两大功能之一，即身份识别和活动要有情况的反映。这是公民个人信息范围扩大的表现。应当指出，前者的可识别性也应当赋予后者增加的“行踪轨迹”，此行踪轨迹不同于现实一般意义上的个人行踪也不同于在网络中个人浏览记录，应当是指通过行踪轨迹可识别公民个人的身份、家庭住址等包括其他个人隐私在内的个人信息，如此才能表明刑法的严厉性，更好地将行为规制入罪。[8]《民法典》第 1034 条第 2 款对个人信息的界定基本沿用了《网络安全法》，但是，该条第 3 项规定：“有关个人隐私的规定适用于个人信息的个人隐私信息。如果没有规定，则适用有关个人信息保护的条款。”由此可见，民法中的个人信息是个人信息和隐私信息交叉的，而不是泾渭分明，尽管它也是以可识别性作为定义的关键特征。[9] 应当指出，个人信息与个人隐私确实有所区别，但刑法所体现的个人信息应当包含个人隐私。首先我国刑法条文中并未就个人隐私单独规定罪名，仅规定了与之概念相似的侵犯公民个人信息罪，理应个人信息包含个人隐私，不然在刑法中个人隐私脱离于个人信息之外，势必会造成更大的法益侵害，从而不利于社会秩序的稳定。

3 规制恶意爬虫行为之价值导向

每一项新兴技术都会带来法律风险，网络爬虫也不例外。在恶意的爬虫行为中，获取数据本身和途径都具有刑事风险。[10] 刑法是一整个社会和公民的最大安全感来源，有刑法作为底线保障公民的生活秩序，能够带给公民极大的安全感，从而对社会稳定与发展产生极大的作用，是必不可少的因素。社会的发展是必然的，随之带来的风险也是必然的，并且不可否认的是，我国目前已经进入风险社会。网络爬虫的出现使得公民的个人信息、个人隐私、法人的私密信息的安全性得不到有效的保障，如果刑法对此危害行为不予以事先预防，那么公众将会依靠自己有限的防卫力量保护自己的合法权益不受侵害，届时社会秩序将会更加难以保障。

但应当注意的是，刑法作为其他部门法的保障法，同时具有补充性以及严厉性。爬虫行为所产生的风险不应全由刑法予以规制，而仅应对恶意的爬虫行为进行制裁。当爬虫行为是善意时，即便造成了一定的风险，也不应当以刑法入罪。

参考文献：

[1] 孙禹 . 强行爬取公开数据构成犯罪吗 [J]. 国家检察官学报 ,2021,29(06):121-13.
[2] 焦旭鹏 . 现代刑法的风险转向——兼评中国当下的刑法观 [J]. 西南民族大学学报 (人文社科版),2018,39(12):79-90.
[3] 张国滨 . 边沁的功利主义刑法思想 [J]. 商情 (科学教育家),2007,(12):119-120.
[4] 赵辉 . 利用网络爬虫暴力爬取作品数据行为的技术属性再评价——以全国首例“爬虫入刑案”之规制立场为视角 [J]. 山西青年职业学院学报 ,2022,35(01):68-71.
[5] 孙禹 . 论网络爬虫的刑事合规 [J]. 法学杂志 ,2022,43(01):162-172.
[6] 苏桑妮 . 从数据载体到数据信息：数据安全法益本位之回归［J］. 西南政法大学学报 ,2020,22(06):97-108.
[7] 宋行健 . 滥用网络爬虫技术收集个人信息的刑法规制 [J]. 湖南科技大学学报 (社会科学版),2021,24(04):139-148.
[8] 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.
[9] 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(03):59-65.
[10] 邱波 . 滥用爬虫技术的刑事风险与刑法应对 [J]. 信息安全研究 ,2020,6(09):830-838.

Criminalization of Web Crawlers in the Era of Big Data

LI Jia—xiao

(College of Criminal Justice，Shanghai University of Political Science and Law, Shanghai 201710,China)

Abstract: In the era of big data, data information has exploded, and web crawlers have emerged as a data search engine, but in the process of using it, the legal benefits of data security have been threatened. As a neutral technical tool, web crawlers are only regulated by criminal law when they are classified as malicious crawlers, and due to the shortcomings of the law itself, the criminalisation path of malicious crawlers needs to be further clarified. Throughout the practice, the majority of cases involving crawlers are related to the infringement of computer information systems and the infringement of personal information of citizens, while the former involves more crimes of suspected illegal intrusion into computer information systems and illegal acquisition of computer information system data. In general, the criminalisation of malicious crawling cannot be divorced from the basic principles of criminalisation, i.e. the principle of unity of subjectivity and objectivity and the principle of statutory penalties.

Key words:web crawlers；malicious crawler behaviour；the principle of unity of subjectivity and objectivity；the principle of the legality of crime and punishment