

上海检察机关护航数字新经济

预防企业把“益虫”用成“害虫”



“爬” 淘宝内部数据盈利，获刑两年半

在本届论坛提供的一组涉非法使用网络爬虫案例中，记者注意到前不久发生在本市的一起非法获取计算机信息系统数据罪案。2018年至2020年，一彩公司（化名）在未经淘宝公司授权许可的情况下，由被告人李某某决策通过非法手段抓取淘宝直播数据，并通过一彩公司开发的某款小程序出售牟利。在李某的授意下，一彩公司部门负责人被告人王某、高某等人分工合作，以使用IP代理、“X-sign”签名算法等手段突破、绕过淘宝公司的防护机制，再通过数据抓取程序（俗称“爬虫”）大量抓取淘宝公司存储的各主播在淘宝直播时的开播地址、销售额、观看PV、UV等数据。至案发，一彩公司整合非法获取的数据后通过微信小程序对外出售牟利，违法所得共计人民币（以下币种相同）22万余元。近日经徐汇区人民法院提起公诉，本案在徐汇区人民法院开庭审理。

三被告面对检方指控后提出,其涉案手段未侵入淘宝直播系统及影响安全运营,涉案数据不具有商业秘密性、隐私性;涉案数据抓取行为的侵害性较弱,社会危害性较小,抓取数据的核心程度不高,不属于知识产权或隐私权保护性质,是否由刑法规制有待商榷。

法院认为,三被告使用 IP 代理目的是为了绕过淘宝公司对同一 IP 短时间高频次发送请求进行拦截的安全保护措施。“X-sign”算法是服务器验证合法请求的协议,是系统辨别是否正手淘 APP 客户端发出请求的安全保护措施,如不破解该算法就无法伪造成合法请求的数据包,请求会被拒绝。三被告明知其行为有法律风险,淘宝公司对除个人用户正常网络浏览外的“爬取”数据行为予以禁止或限制并使用“反爬虫”机制,仍商议利用技术手段对抗“反爬虫”安全措施,其行为应认定为“侵入”计算机信息系统。

法院指出，一彩公司以淘宝数据为卖点的某小程序产品，内容涉及主播位置、PV、UV、销量、IP地址、销售额、标签等，内含无法通过正常账号登录获取淘宝公司存储的且不对外公开或是限制权限查看数据。随着社会迈入数据时代，数据作为生产要素、核心竞争力之一，具有相应商业、社会价值，具有法律保护的法益。一彩公司非法获取淘宝直播数据后，寻求商业化途径，形成多种数据产品，并通过微信小程序出售牟利，经司法鉴定共计获利22万余元，依法属于情节特别严重。

最终，法院支持了检方的公诉意见，依法判决三人罪名成立，判处三人有期徒刑1年3个月至2年6个月不等，并处罚金1至3万元不等。

互联网“工蜂”为数据收集者提供便利

论坛期间，杨浦区检察院第三检察部副主任罗宾告诉记者：“网络爬虫又称为网络

□法治报记者 夏天

如果互联网世界存在一种“工蜂”，那么“网络爬虫”技术当之无愧。它为数据收集者提供了极大便利，也为网站带来更高点击量，许多网络爬虫都成为数字经济的“益虫”。但与此同时，一些不法分子也混迹于爬虫使用者中，利用该技术“爬”走其他互联网公司的内部数据甚至个人隐私数据，危害计算机信息系统安全，扮演了“害虫”角色。近日，杨浦区人民检察院举办了长三角数据合规论坛（第三期）暨数据爬虫的法律规制研讨会，聚焦司法应如何辨别互联网世界的“益虫”，精准打击“害虫”，也为中小企业提供风险预防合规，以更好护航数字经济。



近日，杨浦区人民检察院举办长三角数据合规论坛（第三期）暨数据爬虫的法律规制研讨会。图为与会嘉宾在论坛期间展开讨论

杨浦区检察院 供图

蜘蛛或网络机器人，是互联网时代一项普遍运用的网络信息搜集技术，是按照一定规则自动抓取网络信息、数据的程序或者脚本，能够帮助企业或者个人更便捷、有针对性地获取网络数据信息，目前已经被广泛应用于各种互联网商业模式和应用场景。”

而在论坛主题发言人、欧莱雅中国区数字负责人刘煜晨看来,“网络爬虫就是模拟人类通过浏览器或手机应用上网,让其高效地去网上抓取它所需要的信息这一过程。”

罗宾对记者指出,网络爬虫技术为数据收集者提供了极大的便利,也给专业网络爬虫公司带来巨大的收益,“被爬”的网站也增加了点击量,这属于多赢局面。随着数据资源的爆炸式增长,网络爬虫的应用场景和商业模式变得更加广泛和多样,较为常见的

有新闻平台的内容汇聚和生成、电子商务平台的价格对比功能、基于气象数据的天气预报应用等等。网络爬虫作为数据抓取的实践工具,构成了互联网开放和信息资源共享理念的基石,如同互联网世界的一群“工蜂”,不断地推动互联网的建设和发展。



 网络爬虫有“益虫”
也有“害虫”

“如果说遵守互联网行业规范的爬虫可视其为‘益虫’；反之，那些会分析并自行构造参数对非公开接口进行数据爬取或访问，获取对方不愿意提供的数据的‘害虫’，则常被作

为新型网络犯罪的手段，进而触犯刑法。”罗宾指出，恶意网络爬虫行为一是涉嫌扰乱计算机信息系统正常运行；二是涉嫌侵入、控制或破坏计算机信息系统，危害计算机信息系统安全；三是涉嫌造成相关数据泄露，如利用恶意爬虫破解技术防范措施，非法获取公民个人信息、商业秘密、国家秘密等。

刘煜晨也指出，爬虫如果滥用，就可能干扰网站的正常运营，过多的访问请求易导致服务器崩溃，影响网站的正常运营，对方需要花更多钱以及部署更多维护工程师，来确保服务器能接受更多请求，不至于因崩溃而影响正常经营。此外，爬虫爬到不该爬的信息，涉及对方商业机密，就容易引发更严重的后果。

新浪集团诉讼总监张喆在论坛主题发言中表示,“抓取+展示、抓取+售卖,是数据不正当竞争行为的两大种类。”她表示,抓取+展示类包括未经加工的平台内容数据,用户动态数据和经过筛选加工的平台数据,而抓取+售卖类则包括平台内容数据、平台用户关系数据和平台账号信息数据。

[相关链接 >>>](#)

规制趋势已现 为数字经济“未雨绸缪”

因为爬虫“益虫性”“中立性”的存在,也成为部分“害虫”浑水摸鱼的障眼法。在多年诉讼实践中,张■也总结出被告人的一系列抗辩理由,多集中在“技术中立”即爬虫技术并不具有非法性、抓取数据性质为公开和无需登录即可获得、为了产品及商业模式创新、行为有限且合理不具替代性、反不正当竞争法第二条适用应当保持谦抑性等。

在民事领域,针对爬虫规制的趋势已经呈现。华东政法大学高富平教授认为,整体数据资源的财产性价值将得以确认,将平台控制的整体数据作为一个整体予以保护,一方面摆脱了以往针对具体形式数据的法律适用问题的困境与争议,如是否适用版权法,另一方面也从技术上划定了网络平台数据财产的边界,即以相对确定的网络平台构架为划分依据以确定具有流动性的数据财产权客体范围。

在刑事领域，罗宾表示，我国已有法律对网络爬虫进行规制主要集中在刑法有关计算机信息系统犯罪的相关条文上。从刑法所追求的法益来看，刑法规范的是对目标网站造成严重影响并具有社会危害性的数据抓取行为。若行为违反刑法的相关规定，通过网络爬虫访问收集一般网站所存储、处理或传输的数据，可能构成刑法中的非法获取计算机信息系统数据罪；如果在数据抓取过程中实施了非法控制行为，可能构成非法控制计算机信息系统罪。此外，由于使用网络爬虫造成对目标网站的功能干扰，导致其访问流量增大、系统响应变缓，影响正常运营的，也可能构成破坏计算机信息系统罪。此外，《网络安全法》没有对爬虫行为作出明确规定，但是其司法解释写道“未经授权爬取用户手机通讯录超过50条记录；未经授权抓取用户淘宝交易记录超过500条；未经授权爬取用户运营商网站通话记录超过5000条；未经授权读取用户公积金社保记录的超过50000条的”，可以入刑。

市检察院第四检察部检察官翁音韵告诉记者,由于刑法的谦抑性,其只能在网络爬虫行为产生严重社会危害而无刑罚以外手段进行规制的情形下起到惩治效果,而对于网络爬虫妨碍其他网站正常运行、过量访问收集数据等一般性危害行为很难起到规制作用,因此我国需要建立在刑法以外的行政规制手段,构建完善的刑事责任、行政责任乃至民事责任体系,以保护互联网平台的合法权益,维护网络空间的正常秩序。

翁音韵还表示，精准打击利用网络爬虫犯罪是检察机关的一重职能，除此之外，参与填补这项法律空白，在数字经济时代为中小企业提供网络爬虫应用的刑事合规保障，以预防企业把“益虫”用成“害虫”，最终能够合法合规地“把生意做成”，则是上海检察机关维护网络安全、护航在线新经济的另一重职能。