

数据爬取还有生存空间吗?论新《反不正当竞争法》的爬虫条款

作者: 杨迅

2025年6月27日,全国人民代表大会常务委员会通过了《中华人民共和国反不正当竞争法》(“《反不正当竞争法》”)修订草案。新修订的《反不正当竞争法》将于2025年10月15日起施行。与现行的《反不正当竞争法》相比,新法增加了第十三条第三款,规定经营者不得“以欺诈、胁迫、避开或者破坏技术管理措施等不正当方式,获取、使用其他经营者合法持有的数据,损害其他经营者的合法权益,扰乱市场竞争秩序”。这一规定直指当前互联网生态中盛行的数据爬取行为。

那么,有了这样一条禁止性规定,数据的爬取是否仍然可行?数据爬取的空间又在哪里呢?

一. 爬虫条款的前世今生

网络爬虫,简称爬虫,是指通过编写程序,按照一定规则自动地从互联网上获取信息的过程。这些程序会像在网站之间穿梭的蜘蛛一样访问网页,提取所需的内容,包括文本、图片、视频等数据,继而开展存储或者后续分析等操作。网络爬虫并非黑客行为,通常并不需要侵入计算机的内部,而是像浏览者一样浏览网页本身。然而,与自然人浏览者相比,它具有极高的效率,能够在短时间内抓取大量的数据,因此,网络爬虫是一种高效的,获取第三方网站上已经存在数据的工具。

关于爬虫技术的合法性,一直是一个饱受争议的话题。一方面,它有助于打破数据孤岛,汇聚数据,产生价值;另一方面,它又可能成为某些人不劳而获的手段。我国曾多次试图以立法形式规制网络爬虫行为,但由于爬虫技术的复杂性,多次努力均无功而返。

.....
如您需要了解我们的出版物,
请联系:

Publication@llinkslaw.com

2021年8月，市场监管总局颁布了《禁止网络不正当竞争行为规定》征求意见稿。其中，第二十条规定，“经营者不得利用技术手段，非法抓取、使用其他经营者的数据，并对其他经营者合法提供的网络产品或者服务的主要内容或者部分内容构成实质性替代，或者不合理增加其他经营者的运营成本，减损其他经营者用户数据的安全性，妨碍、破坏其他经营者合法提供的网络产品或者服务的正常运行”。这条规定是在立法上禁止非法爬取数据的一次有益尝试：它试图从爬虫行为的后果着手，界定非法的数据爬取行为，将那些对被爬取内容构成实质性替代、影响数据安全、妨碍其他经营者网络产品和服务正常运行的爬虫行为认定为非法。但是，该条款未能就爬取行为本身的适当性展开分析，在法律未确认“数据”所有权的情况下，也未能就何为“其他经营者的数据”作出界定。

在2024年5月颁布的《网络反不正当竞争暂行规定》正式稿中，前述条款被简化为：“经营者不得利用技术手段，非法获取、使用其他经营者合法持有的数据，妨碍、破坏其他经营者合法提供的网络产品或者服务的正常运行，扰乱市场公平竞争秩序”。有关“实质性替代”、“减损数据安全性”、“增加运营成本”等损害结果被删除，这使得修改后的条款适用范围被大大限制了。这很大程度上反映了立法界对需要立法禁止的爬虫行为还缺乏统一认识。

2024年12月，全国人大常委会公布了《反不正当竞争法》修订草案，向社会公众征求意见。该草案第十三条第二款第四项禁止“以欺诈、胁迫、电子侵入等不正当方式，获取并使用其他经营者合法持有的数据”。这条规定的禁止非常之广：不论后果如何，只要爬虫入侵了计算机系统，获取他人合法持有的数据，即可被认定为不正当竞争行为。不过，这个条款存在两个不确定性：

其一，如何认定“电子侵入”。爬取数据通常只需要接触第三方网站公开部分，虽然也需要破解一些反爬手段，但与入侵计算机内部系统是截然不同的。

其二，如何认定其他经营者合法持有的数据。“持有”一词并不是一个严格意义上的法律概念。纵观《民法典》，仅仅在仓单、提单、上市许可等情境下使用“持有”一词，而就民事权益而言，只有所有权(包括占有、使用、收益、处分)、担保物权、用益物权之说，而没有对“持有”做出界定。

在最终颁布的《反不正当竞争法》中，该条款再次被修正为：“经营者不得以欺诈、胁迫、避开或者破坏技术管理措施等不正当方式，获取、使用其他经营者合法持有的数据，损害其他经营者的合法权益，扰乱市场竞争秩序”。这一规定综合了过去有关爬虫行为的各法律条文，包括草案或正式规定之所长，从爬取的行为、被爬数据的性质及对其他经营者权益和市场竞争秩序的影响等多个方面试图界定立法禁止的爬虫行为。

那么，在新的反不正当竞争法的爬虫条款下，数据爬取行为还有生存的空间吗？

二. 技术管理手段

在新的《反不正当竞争法》下，爬虫行为构成侵权而被禁止，需要满足以下三个条件：第一，从行为上看，避开或者破坏技术管理措施；第二，从目的上看，获取、使用其他经营者合法持有的数据；第三，从结果上看，损害其他经营者的合法权益，扰乱市场竞争秩序。

接下来，本文分别展开分析这三个要件。

首先，何为避开或者破坏技术管理措施？如果被爬数据的经营者通过技术管理手段限制对数据的访问，比如部分数据本来属于非公众普遍所见状态，那么行为人通过技术手段避开或者突破这些限制，从而获取受限制访问的数据，这种情况下当然侵犯了被爬数据经营者的权利。不过，由于这种受限制访问的数据一般具有商业秘密的属性，爬取行为本质上属于《反不正当竞争法》第十条第一款第(一)项规定的“以电子侵入方式……获取权利人商业秘密”的行为，并非本爬虫条款所管制的对象。

那么，何为本爬虫条款所说的“避开或者破坏技术管理措施”呢？是指避开或者破坏反爬措施吗？如果是指避开或破坏反爬措施，怎样的“避开或者破坏”行为具有非法性？

例如，如果反爬措施是限定在一定时间内访问次数，那么行为人将数据爬取行为的次数限定在被爬数据经营者限定的频率范围内，是否属于避开或者破坏行为呢？

又比如，反爬措施是识别和禁止非自然人行为的访问，那么爬虫模拟自然人的访问行为算不算避开反爬措施呢？爬虫本身的技术特性就是以用户的方式去访问和获取网站上的数据，如果具有更高的仿真性，又有什么过错呢？

在现有的一些判决中，法院甚至认定 Robots 协议也是技术手段之一。且不论 Robots 协议只是网站上的一段 txt 文本，在某种程度上是网站自身是否愿意被爬取数据的一种宣示，不具有当然的技术阻断效力。单方面的宣示能够成为技术管理手段吗？这点还有待商榷。

三. 数据持有权

新《反不正当竞争法》所禁止的爬虫行为，以获取、使用“其他经营者合法持有的数据”为目的。那么，如何界定“合法持有的数据”呢？

“持有”并非严格的法律概念，也没有在《民法典》中被界定。《民法典》第一百二十七条规定：“法律对数据、网络虚拟财产的保护有规定的，依照其规定。”但法律没有就“数据”的权利作出规定。

有关数据权益的主要规范性文件是中共中央和国务院于 2022 年 12 月发布的《关于构建数据基础制度更好发挥数据要素作用的意见》，即通称的“《数据二十条》”，其中第二部分第(三)条提及“界定数据生产、流通、使用过程中各参与方享有的合法权利，建立数据资源持有权、数据加工使用权、

数据产品经营权等分置的产权运行机制”，即所谓的“数据三权”。且不论《数据二十条》并非严格意义的法律，只是政策性指引，其中所提及的“数据资源持有人”的含义也有待进一步界定。

有学者认为，数据持有人是一个较为宽泛的概念。比如姚佳教授认为“数据资源持有人是指相关主体基于法定或约定而对数据有权持有的情形。它强调主体对数据的直接控制、持有这样一种状态”并且“持有主体未必享有使用、收益、处分等权能，但从消极意义上讲，其仍然对持有的数据负有安全保障义务”。这一理解，似乎与《数据二十条》的本意是呼应的。一方面，在《数据二十条》下，持有人的对象是“数据资源”，可以是原生态的数据，也可以是公共服务领域自然产生的数据。

但是，这样的理解在《反不正当竞争法》下，是否过宽了呢？目前，在司法实践中，禁止爬取的数据大体分为两类：衍生数据和集合数据。

(1) 衍生数据

衍生数据指的是“在巨量原始网络数据基础上通过一定的算法，经过深度分析过滤、提炼整合以及匿名化脱敏处理后而形成的预测型、指数型、统计型”的数据。在(2018)浙 01 民终 7312 号案件(“**生意参谋案**”)中，法院认为“生意参谋”数据产品中的数据内容“系淘宝公司付出了人力、物力、财力，经过长期经营积累而形成，具有显著的即时性、实用性，能够为商户店铺运营提供系统的大数据分析服务，帮助商户提高经营水平，进而改善广大消费者的福祉，同时也为淘宝公司带来了可观的商业利益与市场竞争优势。‘生意参谋’数据产品系淘宝公司的劳动成果，其所带来的权益，应当归淘宝公司所享有。”

(2) 集合数据

集合数据指的是“长期经营打造平台积累的”数据，是单项数据的整合。在(2019)浙 8601 民初 1987 号(“**微信群控案**”)中，法院认为微信用户数据，可以分为两种数据形态：数据资源整体和单一数据个体。两者对应的是不同的数据权益：微信平台数据就数据资源整体概念而言，腾讯享有竞争性权益；就单一数据而言，腾讯仅享有有限使用权。

无论是衍生数据还是集合数据，都是相关经营者通过创造性劳动抑或通过长期的投资、经营和积累形成的数据资产，而并非原始数据资源。套用《数据二十条》的概念，经营者都拥有数据资产的经营权，而非简单地对原始数据的持有。那么新《反不正当竞争法》下的爬虫条款，是否是对这一范围的突破呢？是否任何数据，包括原始的、零散的数据爬取都属于获取“经营者合法持有”的数据呢？

四. 损害合法权益、扰乱市场秩序

相比修订草案，新《反不正当竞争法》的爬虫条款增添了后果性要件，明确：被禁止的爬虫行为必须是“损害其他经营者的合法权益，扰乱市场竞争秩序”。这一增添无疑是非常重要的，即：数据爬取行为不必然是不正当竞争，除非其带来了损害其他经营者合法权益或者扰乱市场竞争秩序的后果。虽

然,这一后果性的描述仍然比较模糊,但是这种模糊性本身反映了《反不正当竞争法》试图规范变化多端的市场行为的特点。

从目前的司法实践看,被认定不正当竞争的后果有以下几类:

(1) 实质性替代

与被爬数据经营者构成不正当竞争的基本逻辑在于:如果利用其他经营者的数据,与其开展类似的、可替代的和竞争性活动,是一种“搭便车”的行为。被爬数据的持有人的数据虽然可以被共享,但是如果行为人只对数据进行明显替代性或同质化的利用或者对于流量的简单分流,并未对数据进行深度开发、衍生,仅仅是“搬运+快速变现”,则不利于消费者权益的增加、良性的竞争秩序和数据经济的整体发展,属于不正当竞争行为。

(2) 破坏网络生态

破坏网络生态指的是虽然不与被爬数据经营者直接竞争,但是其运营破坏了原有网络产品或者服务设置的规则或者平台的安全运行或评估机制,而被爬数据经营者的利益是基于该等平台安全运行和评估机制的。法院在具体案件中考察的因素包括:具体爬取数据的方式是否突破IP访问限制、封禁措施、反爬措施等;爬取的量级、频率、时间、服务器负荷等是否破坏原有网络产品或者服务的正常运行或者平台的安全运行;爬取行为在效果上是否妨碍原有网络产品或者服务设置的规则或者平台的运行、评估机制,例如数据展示的方式和规则、依据真实访问数据评估公众号文章的机制、产品基于社交而非营销的生态定位等。

(3) 破坏竞争环境

破坏竞争环境更多是从长远角度对市场竞争秩序的破坏和对消费者总体福祉的减损。其包含两个考察要点:

其一,是否违反法律法规和诚实信用的商业道德。人民法院通常结合案件具体情况,综合考虑行业规则或者商业惯例、经营者的主观状态、交易相对人的选择意愿、对消费者权益、市场竞争秩序、社会公共利益的影响等因素,依法判断数据爬取行为是否违反商业道德。在司法实践中,侵犯用户知识产权、人身权或者违反网络安全管理要求的数据爬取,往往被认为是违反“法律法规和诚实信用的商业道德”的。

其二,对消费者/用户长远福祉的影响。法院可能考察平衡各方权益(社会公共利益、经营者权益、消费者权益、市场竞争秩序)和鼓励技术创新,以期获得最优化的公共福祉。在具体案件中,法院往往会综合判断数据爬取行为是否对被爬数据开展深度挖掘、创新,有无更深层次的应用,是否提升社会整体公共利益,从而综合判断数据爬取行为对消费者/用户长远福祉的影响。

五. 结语

在《反不正当竞争法》的框架下，爬虫条款虽对某些不正当的爬虫行为予以禁止，但并未对数据爬取行为进行一刀切的全面封禁。在实际操作中，爬虫行为是否合法合规，往往需要结合具体情境进行个案分析。这就要求相关从业者和企业利用爬虫技术获取数据时，必须充分考量法律的边界与伦理的要求，确保自身行为在合法合理的范畴之内，从而在数据驱动的商业竞争中实现健康、有序的发展。

如您希望就相关问题进一步交流, 请联系:



杨 迅

+86 21 3135 8799

xun.yang@llinkslaw.com

如您希望就其他问题进一步交流或有其他业务咨询需求, 请随时与我们联系: master@llinkslaw.com

上海

上海市银城中路 68 号
时代金融中心 19 楼
T: +86 21 3135 8666
F: +86 21 3135 8600

北京

北京市朝阳区光华东里 8 号
中海广场中楼 30 层
T: +86 10 5081 3888
F: +86 10 5081 3866

深圳

深圳市南山区科苑南路 2666 号
中国华润大厦 18 楼
T: +86 755 3391 7666
F: +86 755 3391 7668

香港

香港中环遮打道 18 号
历山大厦 32 楼 3201 室
T: +852 2592 1978
F: +852 2868 0883

伦敦

1/F, 3 More London Riverside
London SE1 2RE
T: +44 (0)20 3283 4337
D: +44 (0)20 3283 4323



www.llinkslaw.com



Wechat: Llinkslaw

本土化资源 国际化视野

免责声明:

本出版物仅供一般性参考, 并无意提供任何法律或其他建议。我们明示不对任何依赖本出版物的任何内容而采取或不采取行动所导致的后果承担责任。我们保留所有对本出版物的权利。

© 通力律师事务所 2025