

# 数据爬取的不正当竞争认定规则研究

周辉平\*

---

**摘要:**平台对平台内数据享有什么权益是判断数据爬取行为正当性的前提问题。平台对不同形态的数据投入的劳动和资金具有显著的不同,应进行分类保护。用户直接生成展示在网页上的数据,具有原始性和公开性的特征,应采用竞争性权益的保护方式;而平台进行适度处理可供商业化利用并采用密码等保护措施仅向特定人提供的数据,可采用财产性权益的保护路径,对数据爬取不正当竞争行为的认定应由此而展开。司法实践中形成的“三重授权”规则是财产性权益的保护方式,而“实质性替代”和“破坏性利用”规则是竞争性权益的保护方式,我国在构建数据不正当竞争的法律规则时,应总结司法经验并对其进行优化和完善。

**关键词:**数据爬取;不正当竞争;数据权益

---

DOI:10.13519/b.cnki.nulr.2023.02.011

2022年12月2日,中共中央、国务院发布《关于构建数据基础制度更好发挥数据要素作用的意见》,提出“探索有利于数据安全保护、有效利用、合规流通的产权制度和市场体系。”当下《反不正当竞争法》第三次修订正在紧锣密鼓地进行之中,数据不正当竞争条款是此次修订的重点和亮点。在大数据时代的背景下,数据权益保护日益成为学术研究的热点,学界涌现了大量有关数据爬取和数据不正当竞争行为的研究,但对于具体的数据利用准则仍未有相对一致的看法,最新发布的《反不正当竞争法(修订草案征求意见稿)》的数据竞争条款也引发了较多的争议。本文通过总结归纳近年来司法实践中形成的裁判规则,尝试在反思、批判的基础上提出数据不正当竞争的认定方法,以期为法律规则的构建提供参考。

## 一、裁判规则的演化与争议

数据不正当竞争裁判规则的发展与数据经济价值的变化和 data 竞争形势的演变密切相关。纵观我国对数据爬取行为的司法裁判历史,可以发现不同时期规制重点有所不同,由此演化出几种不同的裁判规则。

---

\* 南京农业大学人文与社会发展学院教授,法学博士。本文是笔者主持的国家社科基金“数字经济背景下企业数据权属及利用规则研究”(20BFX122),以及教育部人文社会科学一般课题“大数据利用的竞争法规制研究”(19YJA820062)的部分研究成果。

### (一) 数据的产品市场阶段：“实质性替代”规则

在数据市场早期,数据的利用方式主要是原生性使用,即将数据作为产品直接提供给大众,满足人们的各种需求,例如媒体网站提供新闻信息,生活网站提供消费、娱乐等信息,影视网站提供影视剧作品,我们可以称之为数据的产品市场阶段。在此阶段经营者之间的数据竞争主要表现为同类数据产品经营者之间因盗用网站内容而产生的横向竞争。从2011年大众点评诉爱帮网<sup>[1]</sup>开始,出现了一系列数据抓取的不正当竞争案件,如大众点评诉百度、<sup>[2]</sup>百度诉奇虎、<sup>[3]</sup>腾讯诉头条、<sup>[4]</sup>抖音诉刷宝 app<sup>[5]</sup>等,其主要的争议是竞争者抓取他人经营的网站上的文字、图片、视频等数字资源作为自己的网站内容呈现给大众,由此引发的不正当竞争案件。在这些案件中,法院确立了“实质性替代”标准,认为“利用他人网站的信息不得造成对该网站的市场替代”。

法院认为被告利用技术手段抓取他人网站上的资源在自己的网站上直接展示并获取商业利益,是典型的“不劳而获”和“搭便车”的行为,但并没有采用“未经许可利用他人数据资源必然是不正当竞争行为”的逻辑,而是将“实质性替代”作为合法与非法抓取的行为边界,反映了对数据控制方的财产投入、数据使用者的竞争自由,以及公众自由获取信息的利益的综合平衡。“市场替代标准”最早在2011年大众点评诉爱帮网案中提出,法院指出:“使用垂直搜索技术的网站对于特定行业网站的信息的利用,应控制在合理的范围内,不得对该网站造成市场替代的后果。”超出合理范围是指“已达到了网络用户无须进入大众点评网即可获得足够信息的程度”。<sup>[6]</sup> 该案之后,2016年大众点评诉百度案中进一步完善了该标准,指出在判断行为性质时应综合考虑以下几个因素:1. 利用者的行为是否具有积极的效果,是否有利于商业模式创新、提升用户体验或丰富消费者的选择;2. 使用涉案信息是否超出了必要的限度,理想状态下利用他人资源应当遵循“最少、必要”的原则;3. 行为对市场秩序所产生的影响,是否可能使得其他市场主体不愿再就信息的收集进行投入。<sup>[7]</sup> 在后续类似案件中,“实质性替代标准”相继得到贯彻,成为一个较为成熟的裁判规则。

### (二) 数据的要素市场阶段：“破坏性利用”和“三重授权”规则

当大数据和人工智能时代来临之后,数据为机器学习(算法训练)提供原料,其利用方式从原生性使用变成衍生性使用,数据竞争也从同类产品经营者之间提供竞品的横向竞争转化为上下游经营者之间争夺生产资料的纵向竞争。进入到要素市场阶段之后,数据利用者爬取数据的目的可能并不是提供一个与原告类似的产品,而是利用原告的数据创造出一个新的数据产品提供给用户,也即开发生产品,创造数据的再生价值。由于这个新的数据产品与原告产品存在性质和功能上的不同,实质性替代规则不再能够适用,例如,新浪微博诉蚁坊案,被告通过采集原告微博平台数据,形成网络舆情监测工具“鹰击系统”;<sup>[8]</sup>腾讯诉斯氏案,被

[1] 参见北京市第一中级人民法院(2011)一中民终字第7512号民事判决书。

[2] 参见上海知识产权法院(2016)沪73民终242号民事判决书。

[3] 参见北京市高级人民法院(2017)京民终487号民事判决书。

[4] 参见成都市中级人民法院(2019)川01民初5468号民事判决书。

[5] 参见北京市海淀区人民法院(2019)京0108民初35902号民事裁定书。

[6] 参见北京市第一中级人民法院(2011)一中民终字第7512号民事判决书。

[7] 参见上海知识产权法院(2016)沪73民终242号民事判决书。

[8] 参见北京知识产权法院(2019)京73民终3789号民事判决书。

告通过采集原告微信公众号的数据开发了“极致了”产品,提供公众号数据监控与分析服务。<sup>[9]</sup>此类型案件还有腾讯诉搜道、<sup>[10]</sup>前程无忧诉e成<sup>[11]</sup>等。面对此类案件,法院在认定不正当竞争时,更加侧重于考察获取和利用数据行为的正当性。在此阶段不仅数据的利用方式发生了变化,数据的形态也发生了扩展,用户注册信息、关系链数据、用户操作数据等与个人信息密切相关的信息也成为数据利用者抓取的对象。面对更加复杂的衍生性数据利用行为,司法实践中法律适用更趋向于情境化和回应性,其中“破坏性利用”以及“三重授权”规则渐显雏形,但利益衡量和裁判尺度的把握仍然存在较大的争议。

### 1. “破坏性利用”规则

该规则的依据是《反不正当竞争法》第12条,即当数据利用者爬取数据的手段以及利用数据开发的衍生性产品妨碍和破坏了原告提供的产品和服务的正常运行,则认为构成了不正当竞争行为。在腾讯诉搜道不正当竞争案中法院阐述了“破坏性利用”规则的基本思想,法院指出:“网络经济是共生经济,网络企业所掌握的数据资源更多地具有开放性与共享性,如果其他经营者搭便车式地利用了网络企业所掌握的数据资源开展经营活动,只要不是对他人数据资源破坏性利用或有违法律规定,且能够给消费者带来全新体验的,一般不应被认定为不正当竞争。”<sup>[12]</sup>而“破坏性利用行为”通常指向两个阶段:一是数据爬取行为,二是爬取之后的利用行为。

法院通常将违反平台设置的限制措施而爬取数据的行为认定为具有破坏性。突破平台设置的登录限制、IP频率访问限制、验证码等技术措施,或者违背Robots协议爬取数据的行为会被认定为不正当。<sup>[13]</sup>理由是,这会破坏平台的访问运行机制,平台对这种未经授权的爬虫只能投入更多的成本与其对抗;会对平台服务器造成远超正常用户访问的负担,加大原告的运营成本;会影响原告与用户间协议关于数据处理和安全等条款的履行,也会影响平台对外授权并获得相关收益。爬取之后利用数据的行为如果破坏了原告提供的产品和服务的原貌,也会被认定为具有破坏性。例如在腾讯诉搜道案中,被告在原告产品上以外挂技术嵌套软件,爬取并利用原告平台上的数据并增添了新的功能,法院认为此行为会破坏原告平台的正常运行秩序,影响用户的使用体验。<sup>[14]</sup>

“破坏性利用”规则的出发点是否否定数据权利思维,强调只有“破坏性”的数据利用行为才能被认定为非法,使数据资源具有更多的开放与共享空间。但是将违反平台设置的限制措施而爬取数据的行为一般性地认定为非法,而并不对其造成破坏的真实性以及破坏程度进行考察以及设定标准,则可能导致平台强化对数据的控制权,使平台能够自主决定数据是否可以被他人获取,因此如何认定“破坏性”,仍有进一步探讨的空间。

### 2. “三重授权”规则

平台数据中有一部分来自用户的个人信息,在淘友天下与微梦案中,法院确立了涉及个人信息数据获取的“三重授权”规则。该案中法院认为,第三方通过开放数据接口(Open API)获取用户信息时应坚持“用户授权”+“平台授权”+“用户授权”的三重授权原则。即数

[9] 参见杭州铁路运输法院(2021)浙8601民初309号民事判决书。

[10] 参见杭州铁路运输法院(2019)浙8601民初1987号民事判决书。

[11] 参见上海知识产权法院(2019)沪73民终263号民事判决书。

[12] 杭州铁路运输法院(2019)浙8601民初1987号民事判决书。

[13] 参见北京知识产权法院(2019)京73民终3789号民事判决书;杭州铁路运输法院(2021)浙8601民初309号民事判决书。

[14] 参见杭州铁路运输法院(2019)浙8601民初1987号民事判决书。

据提供方向第三方开放数据的前提是数据提供方取得用户同意,同时,第三方平台在使用用户信息时还应当明确告知用户其使用的目的、方式和范围,再次取得用户的同意。<sup>[15]</sup>

“三重授权”规则始于 Open API 模式,这是一种基于开放平台的数据共享模式:数据需求方与数据提供方签订合同,通过数据方提供的开放平台接口调用数据,这是当下数据交易的常见手段。法院认为,第三方获取数据须经平台同意,既是维护平台核心竞争优势的需要,也是保护用户信息的需要;而第三方获取用户个人信息须经数据主体同意,也是保护个人信息的需要。<sup>[16]</sup>在当下的司法实践中,“三重授权”规则只有在属于用户身份信息的情境下适用,并未适用于用户行为信息。例如在微梦诉淘友天下案中,该规则适用于用户头像、名称、职业信息、教育信息、用户自定义标签,以及脉脉注册用户手机通讯录联系人与新浪微博用户的对应关系,这些都是身份信息。而在淘宝诉美景案中,法院认为网络用户浏览、搜索、收藏、加购、交易等行为痕迹与标签信息属于网络用户非个人信息,<sup>[17]</sup>腾讯诉斯氏案中,用户操作数据(含朋友圈点赞评论、支付等)也被认为属于非个人信息。<sup>[18]</sup>

“三重授权”规则的适用范围并不是很清晰,有学者认为“三重授权”规则有从 Open API 模式向授权爬取模式扩张适用的趋势。<sup>[19]</sup>数据利用者获取数据,除了 Open API 接口之外,还可以通过爬虫抓取网页上的数据,网页上的前端数据一般是公开的,如果“三重授权”规则适用于爬取网页数据,则可能强化平台对数据的控制,使第三方获取数据更加困难。

### (三) 问题和争议

我国司法裁判中形成的裁判规则是在个案的情形中发展起来的,除了实质性替代规则较为成熟之外,其余尚处于发育期,不仅规则本身的适用范围和条件不够明晰,规则之间的适用关系也不够明确,某种情况下甚至存在冲突的可能,现阶段构建数据不正当竞争的法律规则,应从反思和检讨裁判规则,理清各规则的适用条件开始。

当下理论和实践中最具争议性的问题,莫过于互联网平台对平台内数据享有什么权利?是否可以控制平台数据的利用?从我国审判实践来看,“三重授权”规则最接近于数据权利理论,第三方利用数据需平台和数据主体(用户)的三重同意,反映了平台和数据主体对数据的强控制,这种高标准虽有利于提供较强的保护,但也可能限制数据的流通和开发,抑制企业创新和数字经济的发展,并导致企业运营成本和社会成本的增加。<sup>[20]</sup>“实质性替代”和“破坏性利用”规则遵循的是行为规制的路径,以数据利用的后果判断行为的正当性,而不是强调主体对数据的排他性控制,能够给予数据利用者更大的自由空间,但司法实践中对数据爬取行为正当性的评价仍然极大地维护了数据控制方即平台的意志,将违反平台限制性措施爬取数据的行为认定为不正当,客观上保护了平台的数据控制权。大数据时代数据是重要的商业资源,平台有强烈的意愿将全部数据占为己有。例如腾讯在《微信开放平台开发者协议》中声明:微信开放平台运营数据、用户数据等数据的全部权利,均归属腾讯,且是腾讯的商业秘密,

[15] 参见北京知识产权法院(2016)京73民终588号民事判决书。

[16] 参见北京知识产权法院(2016)京73民终588号民事判决书。

[17] 参见杭州铁路运输法院(2017)浙8601民初4034号民事判决书。

[18] 参见杭州铁路运输法院(2019)浙8601民初1987号民事判决书。

[19] 参见陈沛:《数据流通与利用中的“三重授权”原则》,载《上海市经济管理干部学院学报》2020年第1期,第61页。

[20] 徐伟:《企业数据获取“三重授权原则”反思及类型化构建》,载《交大法学》2019年第4期,第24页。



依法属于用户享有的相关权利除外。<sup>[21]</sup> 微博也在使用协议中声明:未经微博运营方事先书面许可,用户不得自行或授权、协助任何第三方非法抓取微博内容。<sup>[22]</sup> 法院普遍承认此类协议的效力,并要求数据获取方应完全遵守该协议,这就使我国有关平台数据的利用,具有明显的保护平台控制权的倾向。

但数字经济的繁荣离不开数据的自由流动和充分利用,允许平台控制数据可能造成数据的垄断和不公平竞争,使初创企业无法自由获得数据,从而损害创新。美国第九巡回法院在 *HiQ v. LinkedIn* 案中拒绝将违反平台限制性措施爬取数据的行为认定为非法,并指出如果像 LinkedIn 这样拥有大量公开数据的公司,被允许选择性地禁止潜在的竞争对手访问和使用这些公开数据,那么最初的创新者将无法汇集和分析这些数据。<sup>[23]</sup> 也有学者指出,大公司正在利用《计算机欺诈和滥用法案》(CFAA),试图把该法令变成数据盗用法,这样他们就可以在法律的名义下实施反竞争行为,限制竞争对手获得在网上已经向世界公开的信息。<sup>[24]</sup> 互联网上汇集了大量的数据,是创新和财富的源泉,也是各方争夺的焦点,平台是否有权控制数据的问题越来越突出,这也决定了数据爬取行为正当性的认定。

## 二、平台的数据权益与控制权的合法性

对企业数据采用权利保护还是行为规制一直是近年来学术争议的焦点,但从最近的学术成果来看,两种观点有渐趋融合的趋势:即便是主张权利保护的学者,亦强调要贯彻平衡理念,对商业数据赋权不能一刀切,而要遵循差异化保护路径:例如对于公开数据集合仅从整体意义上给予有限的财产权保护,非公开数据则可以受到专有权利的保护。<sup>[25]</sup> 或者对数据集合设置有限财产权,对数据产品设置新型财产权。<sup>[26]</sup> 而主张采用反不正当竞争法规制的学者,则提出数据专条保护模式应当立足于权利保护。<sup>[27]</sup> 《反不正当竞争法》虽然是行为规制法,但如果对“不正当”行为采用严格界定的方法,同样能够起到类似于权利保护的效果。我国 2022 年 11 月发布的《反不正当竞争法(修订草案征求意见稿)》第 18 条,将保护对象限定于“经营者依法收集、具有商业价值并采取相应技术管理措施的商业数据”。根据学者的解释,凡是能够宣示权利人主张权利的意愿并且可以被识别的措施,即可成为保护措施,这既包括 ID、密码等认证技术,也包括各种“反爬虫”的限制措施以及 robots 协议等。<sup>[28]</sup> 如果这样理解,平台就可以通过“保护措施”来宣誓“权利”,从而将不符合自己意志的爬取行为宣布为违法,这可能会比有限权利保护的做法更加激进。在反不正当竞争法视角下,如何准确理解

[21] 参见《微信开放平台开发者协议》第 2.7.6 款。

[22] 参见《微博服务使用协议》第 1.3.2 条。

[23] *hiQ Labs, Inc. v. LinkedIn Corporation*, 938 F.3d 985 (2019)。

[24] Jamie L. Williams, *Automation is not “Hacking”: Why Courts Must Reject Attempts to Use the CFAA as an Anti-competitive Sword*, B.U. J. SCI. & TECH. L. Vol. 24.2:ppp, 2018

[25] 参见冯晓青:《知识产权视野下商业数据保护研究》,载《比较法研究》2022 年第 5 期,第 44 页。

[26] 参见姬蕾蕾:《企业数据保护的司法困境与破局之维:类型化确权之路》,载《法学论坛》2022 年第 3 期,第 109 页。

[27] 参见孔祥俊:《论反不正当竞争法“商业数据专条”的建构》,载《东方法学》2022 年第 5 期,第 15 页。

[28] 同前注[28],孔祥俊文,第 27 页。

数据权益,如何在平台的经营自主权与访问、接入的容忍义务之间找到平衡点,是首要解决的难点问题。

### (一) 平台数据权益的有限性及分类保护

在大数据时代,平台内汇集了大量的数据,这些数据是用户在使用平台服务的过程中产生的,一部分来自用户直接提交,另一部分则来自平台记录用户的使用行为,可以说用户的使用行为和平台的收集行为共同促成了平台内数据的生成。从劳动产生所有权的观念来看,平台数据的权益应由用户和平台共同享有,任何人只能在播种之处收获,平台主张的权益应与其投入的劳动、资金及其对数据增量价值贡献相适应。对于平台数据而言,用户可以通过携带权分享数据上的利益,<sup>[29]</sup>但毕竟主张携带权的只是极少数用户。既然无数人贡献了平台数据,每一个人既是数据的创造者,又有可能是数据的利用者,平台数据更应具有开放性和共享性,使更多的人能够从数据的利用中获益。数据的共享与互通不但无损于数据本身的价值,而且更可能发挥数据的规模化优势,并为人工智能产业提供坚实的基础。<sup>[30]</sup>

但即便是在价值层面重视数据共享利益也不能完全无视数据收集者的利益,如果平台的利益得不到合理保护,就会抑制其在数据收集、加工上的投入,助长不劳而获的搭便车行为,损害企业的经营利益,因此平台的数据利益需要得到恰当保护。我国已经有不少学者注意到了企业在不同数据上投入的不同,认为应当采用差别化的保护方式,例如数据产品可以采用权利化保护,原始数据只能采用行为规制的方法。<sup>[31]</sup>也有学者认为对原始数据的收集、清洗、加工后所汇集的是数据集合,对数据集合进行深度加工与处理是数据产品,前者应赋予有限权利,后者则可设新型财产权。<sup>[32]</sup>在这些分类中,原始数据、数据集合以及数据产品的边界并不是很清晰。在平台环境下,用户提交给平台展示在网页上,是平台收集而来但未经进一步加工的数据,是原始数据还是数据集合?平台经过初步的清洗、整理向有需求的第三方提供,但未经算法深度分析的数据流,是数据集合还是数据产品?这些界限并不是很清晰。本文采用“用户生成的网页数据”和“平台限定提供的数据”来区分不同投入所形成的数据,也可以对应司法实践中常见的抓取网页前端数据以及突破保护屏障抓取计算机后端数据两种不正当竞争行为,本文认为对这两种数据应采不同的保护方式,由此对数据爬取行为也应在不同情境之下进行评价。

“用户生成的网页数据”是用户直接提交给平台,或者用户在使用平台过程中生成,直接展示在网页上的数据。此类数据具有原始性、公开性的特征,平台仅享有有限权益。我国现有的数据不正当竞争争议中,大量案件属于对用户直接提交或者行为触发的数据引发的争议,“破坏性利用”或“实质性替代”也是应对此类争议产生的规则,法院认为平台对“数据资源整体享有竞争性权益”。当今掌握大量数据的平台如微博、微信、淘宝、大众点评等并非直接从事数据经营的企业,而是诸如从事信息传播、网购、外卖、物流等企业,这些企业虽然为汇聚

[29] 例如欧盟《数字市场法》规定作为网络守门人的平台有义务应用户和用户授权的第三方的要求,免费提供用户使用平台服务时生成数据的可移植性。

[30] 丁晓东:《数据到底属于谁?——从网络爬虫看平台数据权属与数据保护》,载《华东政法大学学报》2019年第5期,第82页。

[31] 参见毛立琦:《数据产品保护路径探究——基于数据产品利益格局分析》,载《财经法学》2020年第2期;李晓珊:《数据产品的界定和法律保护》,载《法学论坛》2022年第3期,第122页。

[32] 前注[26],姬蕾蕾文,第115页。

客户、吸引流量、管理平台付出了大量的劳动和资金,但其付出是为了主营业务,并且这部分投入也已通过其营业收入以及流量变现获得了回报。因此,即便没有数据产权,也不会缺少生产数据的激励。<sup>[33]</sup> 欧盟的数据库权利也被认为不能对此类对象提供保护,因为数据库特别权利保护的是“寻找数据并将其收集到数据库中的投入”,而不是“用于产生数据内容的投入”。<sup>[34]</sup> 我国法院认为:“原始数据对于社会的价值贡献仍未脱离用户信息所包含的资讯内容。数据采集主体在此过程中虽然付出了一定的劳动,但并未提升用户信息的品质,故数据采集主体仅有权享有其劳动所增加的价值而不是原始数据的全部价值。”<sup>[35]</sup> 此类数据应更多地开放共享,允许第三方数据爬虫合理使用。数据是非竞争非排他的,允许第三方爬取和利用并不会影响平台的经济利益。平台通过对原始数据的加工利用可以获得不菲的收入,已经足以弥补其在数据收集上的投入,并不能要求垄断全部的经济利益。在“用户生成的网页数据”上,平台拒绝第三方爬取的理由只能是爬取行为破坏了平台的正当营业活动,或者窃取了平台的经营果实,为了保护自己的营业利益而拒绝爬取,从这个意义上说,平台享有的是竞争性权益。

“平台限定提供的数据”是平台企业对其所收集的数据进行脱敏、过滤、格式调整、加密、筛选等适度加工之后,采用技术手段进行控制并仅向特定人提供的商业化数据。数据共享互通是大数据时代的价值追求,原始数据需要经过清洗、脱敏并转变成适于再利用的格式才能进行共享。共享数据需要对元数据、数据模型和用于数据存储和处理的算法进行补充投资,以保护用于共享数据存储、处理和访问的信息技术基础设施,总的前期成本和支出可能会非常高。<sup>[36]</sup> 为了激励数据提供者将原始数据商业化,弥补其为数据共享所作的投入,对限定提供数据给与更高程度的保护是必要的。所谓“限定提供”,是数据提供方以 ID、密码以及其他管理方式使数据处于私有状态,只向特定人提供的方式。开放数据接口(Open API)就是常见的“限定提供”,数据持有方和利用方基于合同自愿共享数据,协商确定价格,是如今数据交易的主要方式。“限定提供的数据”有的可能仅对原始数据进行了简单处理以使其适于传输,也有的可能是算法深度开发与系统整合的衍生数据,这些数据的价值会有所不同,但就权益状态而言,数据提供方享有的都是相对独立的财产性权益,类似于商业秘密。平台对于限定提供的数据,一方面可以支配数据的利用并获得收益,另一方面可以排除他人干涉,具有权利的特征,但因其排他性的范围有限,并且负有向特定人例如数据主体的共享义务,称其为“财产性权益”的保护状态更为恰当。对于“平台限定提供的数据”,数据利用者获得数据,必须要经过平台的授权,如果未经匿名化处理,还需取得数据主体的同意。

《日本反不正当竞争法》保护的就是“限定提供数据”,根据该法第 2 条,受保护的数据是以营利为目的提供给特定对象的信息,通过电磁方法(电子方法、电磁方法以及其他不能被人

[33] W. Kerber, *A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis*, [2016] GRUR Int. 989, 997.

[34] Hugenholtz, P. B. (2017). *Data Property in the System of Intellectual Property Law: Welcome Guest or Misfit?* In S. Lohsse, R. Schulze, & D. Staudenmayer (Eds.), *Trading Data in the Digital Economy: Legal Concepts and Tools: Münster Colloquia on EU Law and the Digital Economy III* (pp. 75 - 99). Baden-Baden: Nomos.

[35] 参见杭州铁路运输法院(2019)浙 8601 民初 1987 号民事判决书。

[36] OECD: *Enhancing Access to And Sharing of Data-Reconciling Risks and Benefits for Data Re-use across Societies*, 2019. <http://doi.org/10.1787/276aaca8-en>, 2022 - 7 - 15.

的知觉所识别的方法)被储蓄积累到一定数量,并且被管理的技术或商业信息。<sup>[37]</sup> 数据提供方要有反复、持续提供数据的意思,或者虽尚未实际提供,但已经通过投资、贷款、厂房或者机器设备租赁、雇佣工人、广告等行为客观反映出来,并能被一般人认识。接受数据的一方是付费获得的人,或满足特定资格参加数据共享的团体的人。<sup>[38]</sup> 可以看出,日本反不正当竞争法保护的是可以商业化提供的数据,而并未将尚在网页展示阶段的数据作为保护的主体。我国《反不正当竞争法(修订征求意见稿)》第18条,对“采取相应技术管理措施的商业数据”进行保护,看似与日本反不正当竞争法的保护存在相似性,实则不然。征求意见稿没有明确“技术管理措施”指什么,防止网页爬虫的各种措施以及 robots 协议也可以被理解为“管理措施”,因此该条款既规范违反限制措施爬取网页公开数据的行为,也规范违反管理措施获取限制提供数据的行为。而平台对网页公开数据和限制提供数据享有的权益是不同的,对前者享有竞争性权益,对后者享有财产性权益,对突破管理措施获取数据的行为也应分别评价。

## (二) 竞争性权益、财产性权益与数据控制权

平台对“用户生成的网页数据”享有竞争性权益,对“限定提供数据”享有财产性权益,理解这两种权益的不同,是解释平台对数据是否享有拒绝抓取的控制权的关键。

竞争性权益保护的并不是对数据财产的支配和控制,而是正当营业利益不受不正当的数据爬取行为损害,这是一种反向保护的消极权益。竞争性权益不是一种绝对权,也不是财产拥有的状态,而只是通过正当竞争获得盈利的机会。当个别经营者以不正当的方式参与竞争时,其他诚实经营者的利益,最终均表现为现实利润的减少、长远营利能力受损害或者积累的竞争优势被破坏,而这些可以综合表述为竞争利益。<sup>[39]</sup> 数据爬取行为对平台竞争性权益的损害,可以表现为破坏了平台的正当运行,也可以表现为因数据爬取替代了被爬取方提供的产品而导致的营业利益损害,还可以表现为其他竞争利益受损的状态。在市场竞争中,竞争者之间的损害是常态,正当竞争也会损害竞争者利益,因此衡量竞争性权益损害必须坚持整体竞争观的视角,综合衡量竞争者、消费者、其他市场参与者的利益。德国判例曾以对竞争自由构成“显著损害”作为认定门槛,“显著损害”就是“导致竞争者任凭自己的努力都无法将自己的业绩在市场上合适地展示”。<sup>[40]</sup> “显著损害”的要件可以避免竞争性权益绝对化。

财产性权益是竞争性权益向法定权利过渡的中间状态,有学者称之为“应然权利”,是一种十分接近法定权利状态的法益,虽未为法律所确认,却实际构成了法定权利的价值基础。<sup>[41]</sup> 反不正当竞争法主要保护竞争性权益,也保护一部分财产性权益,例如商业秘密。有学者认为,根据现代财产权概念,只要管制或排他的内容可以被概括为权利人排除公众对某种客体的某种利用行为,权利人可以自由转让这一利益,则无论排他权的范围宽窄,其实都可

[37] 刘影、睦纪刚:《日本大数据立法增设“限定提供数据”条款及其对我国的启示》,载《知识产权》2019年第4期。

[38] 李扬:《日本保护数据的不正当竞争法模式及其检视》,载《政法论丛》2021年第4期,第70页。

[39] 参见李友根:《经营者公平竞争权初论——基于判例的整理与研究》,载《南京大学学报(哲学·人文科学·社会科学)》2009年第4期,第63页。

[40] BGH Zulässigkeit von Fernseh-Werbeblockern-Werbeblocker, NJW 2004, 3032.

[41] 参见龙俊:《反不正当竞争法“权利”与“利益”双重客体保护新论》,载《中外法学》2022年第1期,第72页。



以视为一种财产权利。<sup>[42]</sup>从这个角度上说,“限定提供数据”也可以称为一种“权利”,但在反不正当竞争法的视角之下,对其保护仍要考虑竞争者、消费者、其他市场参与者的利益平衡,保护范围具有相当程度的限定性,因此本文仍称之为“财产性权益”。与竞争性权益相比较,财产性权益是一种具有支配和归属效能的主观性权益,具有类似于权利的外在表征,可以支配该数据获得收益,并在特定的范围内排除他人干涉,具有正向保护的积极权能。将“限定提供数据”作为财产权益加以保护,宜由立法对其排他性的范围和数据共享义务加以规定,以限定其保护范围,促进数据的共享和利用。

通过上文分析可以知道,平台仅对“限定提供数据”享有控制权,可以在权限范围内自主决定数据的利用;而对于“用户生成的网页数据”并不享有控制权,不能以自己的意志决定数据是否被第三方抓取,第三方违反 robots 协议,或者突破平台设置的技术措施爬取网页数据,并不当然构成违法,是否构成不正当竞争行为还需进一步论证爬取行为对市场竞争的损害。

### 三、数据不正当竞争行为的认定

平台对“限定提供数据”和“用户生成的网页数据”享有不同的权益,对数据爬取行为构成不正当竞争也应分别评价。

#### (一) 获取和利用“限定提供数据”构成不正当竞争行为的认定

“限定提供的数据”是平台采用技术手段控制并仅向特定人提供的商业化数据。反不正当竞争法保护“限定提供数据”,首先要对其保护的条件的规定,然后对何种行为构成不正当竞争进行规定。

“限定提供数据”的原始来源可能是用户提交的数据,也可能是系统或设备记录的数据,无论何种来源,都需要经过数据控制方的适当处理,将数据转变为可共享的格式之后才可提供给第三方进行商业化利用,尤其是个人数据需要处理避免损害用户的隐私利益,这些方面付出的投入是取得财产性权益的基础。因此,构成“限定提供数据”要符合以下条件:1. 经过适当的处理可以商业化利用;2. 采用密码等管理措施使其处于控制之下;3. 仅向特定人提供,一般情况下需有偿获得。例如 Open API(开放应用程序接口)是服务型网站常见的一种限定提供数据的方法,网站的服务商将自己的网站服务封装成一系列 API 开放出去,供第三方开发者使用,这种开发模式能够更好地发挥数据资源价值,实现开放平台方和应用方之间的合作共赢。对于“限定提供数据”,存在以下行为可认定构成不正当竞争:

##### 1. 未经授权获取数据的行为

数据控制方对“限定提供数据”采用密码等管理措施进行控制,第三方要获取数据,必须取得数据控制方的授权,通过用户 ID、密码等由数据接口获取数据。没有权限的人通过盗窃、欺骗等方式取得登录资格获取数据,或者破坏加密措施而获取数据的,属于违背“诚实信用的商业道德”的数据获取行为。此种行为的不正当性首先体现在无偿侵占了被爬取方的财产性利益。未经授权获取限定提供数据的行为会无偿占有被爬取方所应获得的经济利益。

[42] 崔国斌:《大数据有限排他权的基础理论》,载《法学研究》2019年第5期,第21—22页。

例如在淘友天下与微梦案中,微梦公司所经营的新浪微博通过移动应用开发平台提供相关接口,实现第三方 WAP 站和客户端等多种应用的接入,淘友天下公司在其接口权限到期之后,仍利用微博方权限管理的漏洞获取数据,此种行为损害了微博方的财产性利益。<sup>[43]</sup>

破坏密码等安全管理措施侵入计算机服务器获取数据,还会对计算机信息系统安全形成威胁,其带来的法律后果还可能是刑事责任。例如在谷米公司与元光公司案中,元光公司的工作人员在破解谷米公司“酷米客”客户端的加密算法没有成功之后,聘请技术人员将谷米公司 APP 中的加密系统攻破,进入权利人服务器后台获取谷米公司整合形成的实时公交信息数据,该案中元光公司不仅因不正当竞争行为承担民事责任,还因非法侵入计算机信息系统罪被判决承担刑事责任。<sup>[44]</sup>

未经授权获取“限定提供数据”还可能对用户的隐私利益造成损害。“限定提供数据”可能来自用户,平台有义务保护用户的隐私利益。例如新浪微博 Open API 的权限体系有三个维度:授权有效期限、调用频次权限、接口权限,每个接口都有普通和高级两种权限级别,高级接口权限包括用户的隐私信息,如:职业信息接口,教育信息接口等,高级接口需要单独进行申请,并通过平台的严格审核后才能调用。应用方在获取相关高级接口权限后,调取用户的隐私信息时还需要经用户授权。因此第三方获取隐私信息需平台授权+用户授权,未经授权会同时侵害平台方和用户方的利益。

## 2. 披露和使用未经授权获取的数据

通过上述手段不正当获取“限定提供数据”之后,再进行披露和使用的行为,构成不正当竞争。对于披露和使用行为,关键是要看使用的数据量的多少,仅仅少量的使用尚不能构成不正当竞争。根据日本的经验,使用受保护数据的全部和其中具有利用价值的部分数据;每次获取一小部分,连续或者断续获取的结果,整体上达到了相当数量的情况的,该系列行为应当一体评价为获取的不正当行为。<sup>[45]</sup>

### (二) 爬取和利用“用户生成的网页数据”构成不正当竞争行为的认定

“用户生成的网页数据”是由用户提交给平台发布在网页上的数据,或者用户在使用平台过程中生成的数据。此类数据大量存在于各类平台的网页中,如微博文章、微信公众号发文、大众点评、用户点赞、评论等,这是一种仍处于原始形态未经商业化利用处理的数据,平台对此类数据享有竞争性权益,有权保护平台服务不受不当的数据爬取行为损害。

对“用户生成的网页数据”争议最大的问题是破坏平台设置的保护措施,包括 robots 协议、IP 地址阻碍、验证码等反爬措施是否可直接认定为不正当竞争。前文已经分析过,平台对“用户直接生成的数据”没有排他性的主观权利,并不能以平台意志来决定数据是否可以被爬取。但是平台是否可以以避免爬虫破坏网站的正常运行为由,设定 robots 协议和反爬措施呢?

爬虫作为一种数据收集的技术本身是中立的,并不是所有的爬虫行为都具有破坏性。常见的反爬虫技术措施如 IP 限制、验证码、登录限制等,针对的是网络上静态数据,采取的一般

[43] 北京知识产权法院(2016)京 73 民终 588 号民事判决书。

[44] 广东省深圳市中级人民法院(2017)粤 03 民初 822 号民事判决书。

[45] 参见日本通产省于 2019 年 1 月 23 日发布的《受保护数据指南》,前注[37],刘影、睦纪纲文,第 94 页。

都是前端反爬技术,被爬取的数据基本上是属于公开或者半公开(通过注册程序可获取)的数据,反爬措施只是对某种特定的访问方式进行了限制,并不具有排除特定主体访问的强制性效果,而爬虫也是无法进入被爬虫企业的服务器以及获取任何系统管理权限的。Robots 协议作为一种技术规范,其作用只在于引导爬虫机器人访问,但无论爬虫机器人是否遵守,Robots 协议都不会起到强制禁止访问的作用。因此,Robots 协议、IP 地址阻碍、验证码等措施应该被理解为仅仅是减速带而不是访问限制。<sup>[46]</sup> 只有破解密码系统侵入计算机系统内部获取不公开数据的行为,才具有“侵入计算机信息系统”的破坏性,爬取前端公开数据并不会危害网络安全,也不会侵入计算机系统。Kerr 教授主张密码门可以将网络上的开放空间与封闭空间分隔开,主张密码才是访问限制。<sup>[47]</sup> 美国第九巡回法院也认为只有受密码保护的网站以及其他阻止公众查看信息的网站才需要授权。<sup>[48]</sup> 北京市高级人民法院在 2017 年的百度与奇虎不正当竞争案件中认为,网站除非有正当理由,不得以 Robots 协议拒绝搜索引擎抓取其内容,而正当理由包括:保护受访网站的内部信息或敏感信息的需要;维护受访网站正常运行的需要,如果抓取会导致受访网站无法正常运行,则有必要对其进行限制;出于保护社会公共利益的需要。<sup>[49]</sup> 笔者认为这是对 Robots 协议等限制措施的正确理解和解读。因此,平台设置限制爬取“用户生成的网页数据”的措施应当有正当理由,而对于破坏平台设置的保护措施爬取数据的行为,应采用结果导向的判断方法,看其爬取行为是否造成了实质性损害的后果。

综合司法实践中形成的裁判经验,爬取“用户生成的网页数据”构成不正当竞争,主要有以下三种情况:

#### 1. 爬取和利用数据的行为实质性妨碍和破坏平台的正常运行

网络爬虫是通过模拟人工点击来对网站进行一次性大量访问,这一过程中会对被访问对象的网络资源进行一定的挤占甚至是破坏,网站设置反爬措施的重要理由也是避免过度的爬取行为给服务器带来负担。但是应当注意到,如今随着技术的不断发展,多数网站都已具备负担大规模爬取行为的能力,数据爬取行为仅仅是在突破一定界限的情况下才会对平台运行造成破坏,平台并不能以服务器容量为理由拒绝数据的共享,毕竟服务器能力是会随着技术的发展而不断提升的,“流量问题或数据访问拥堵从本质上而言是一个技术问题,并不是数据公共所有带来的问题”。<sup>[50]</sup>

突破反爬措施爬取数据要达到“实质性”损害的程度才可以被认定为不正当。司法实践中“破坏性利用”规则在适用中存在的问题是没有强调损害的“实质性”,导致可能对爬取行为过度追责。最高人民法院在海带配额案中指出,判断竞争行为不正当的依据只能是经营者的合法权益确实因竞争者的行为受到了实际损害。<sup>[51]</sup> 在数据爬取案中,“实际损害”应体现为损害的显著性。“非显著损害不禁止”等于是赋予了其他竞争者更多的行为自由,避免了将特

[46] Orin S. Kerr, *Norms of Computer Trespass*, 116 COLUM. L. REV. 1143, 1163 (2016).

[47] Orin S. Kerr, *Norms of Computer Trespass*, 116 COLUM. L. REV. 1143, 1170 (2016)

[48] *hiQ Labs, Inc. v. LinkedIn Corporation*, 938 F.3d 985 (2019)

[49] 北京市高级人民法院(2017)京民终 487 号民事判决书。

[50] 参见丁晓东:《论企业数据权益的法律保护——基于数据法律性质的分析》,载《法律科学(西北政法学报)》2020 年第 2 期。

[51] 参见最高人民法院(2009)民申字第 1065 号民事判决书。

定竞争者已经取得的市场利益“权利化”。<sup>[52]</sup>2019年的《数据安全管理办法(征求意见稿)》曾经提出一个爬虫妨碍网站正常运行的参考标准,即:自动化访问收集流量超过网站日均流量三分之一,此时网站可以要求爬取方停止访问。虽然此标准并未在正式稿中得到采纳,但至少说明爬虫访问需要达到一定的破坏性程度才可以被禁止已经得到了一定的共识。<sup>[53]</sup>

从美国法院运用动产侵权理论规制爬虫占用带宽和系统资源的判例发展来看,<sup>[54]</sup>要求证明爬虫导致显著的实质性损害是总体趋势。美国 eBay v. Bidder's Edge 案是早期案例,该案中法院认为,尽管 BE 的行为对 eBay 的带宽和系统资源伤害较少,但如果初步禁令救济被否决,其他公司可能遵循 BE 的例子,那么 eBay 电脑系统的负载将符合实质性损害的标准。<sup>[55]</sup>但此案之后,一些法院对 eBay 案提出的会导致其他爬虫模仿的论点持更加谨慎的态度。2003年,加州最高法院在 Intel v. Hamidi 案中含蓄地反转了 eBay 案的判决,法院认为爬虫模仿理论不是法院的完整观点,也不会是加州或美国一般法律的正确表述,该意见可能不再是有效的先例。<sup>[56]</sup>2015年,在 White Buffalo Ventures LLC v. University of Texas at Austin 案中,第五巡回法庭说:“本院不认为仅仅使用爬虫进入一个公开可用的网站收集信息,而没有更多,就足以满足侵犯动产的损害要求。既然爬虫不会导致动产的物理损害,就必须有证据表明计算机的使用或计算机网络被爬虫的使用不利地影响了。”<sup>[57]</sup>

综上,以妨碍和破坏平台的正常运行为由,要求认定爬取数据的行为构成不正当竞争的,应当有证据证明其损害达到“显著”或“实质性”的程度。

## 2. 爬取和利用数据的行为造成对被爬取方网络产品或服务的实质性替代

禁止不劳而获、食人而肥是不正当竞争理论的重要思想。互联网平台在其提供的产品和服务上投入了劳动、技术和金钱,如果竞争者爬取平台上的数据向用户展示类似的产品和服务,造成对被爬取方产品的替代,会无偿占有被爬取方的竞争优势。司法实践中“实质性替代规则”就是反不正当竞争法中盗用成果理论在数据竞争中的运用。

盗用理论不应被任意扩张,原、被告之间存在直接竞争关系,以及造成实质性替代的后果是数据盗用成立的关键要件,这两个要件可以防止过度保护。只有在原、被告之间存在直接竞争关系的情况下,被告爬取的数据才可能做竞争性的使用。如果爬取方抓取数据之后并未将数据做竞争性使用,而是用作生产新的数据产品,例如将平台上的公开信息抓取之后,进行统计、分析,用来预测行业发展的走向,这种数据再利用并不会替代被爬取方提供的产品和服务,从而也不会无偿占有被爬取方的劳动成果。“实质性替代”要件可以避免过度宽松的适用导致数据权利化。美国在信息盗用的早期判例中,将侵权基础建立在不劳而获的不道德性之上,这种立场因可能导致宽泛的保护而备受争议,后来的判例采纳了一系列的限制性条件,其中要求原被告之间存在直接竞争关系,以及被告行为对原告造成“实质性”损害就是关键的限

[52] 张占江:《不正当竞争行为认定范式的嬗变——从“保护竞争者”到“保护竞争”》,载《中外法学》2019年第1期,第220页。

[53] 参见翟巍、刘一诺:《反不正当竞争法视角下公开数据爬取行为的合法性边界》,载《价格理论与实践》2021年第8期,第70页。

[54] 动产侵权(trespass to chattels doctrine)指的是故意干扰对个人财产的占有,造成了直接损害。

[55] eBay v. Bidder's Edge, 100 F.Supp.2d 1058 (N.D. Cal. 2000).

[56] Intel Corp. v. Hamidi, 30 Cal. 4th 1342, 71 P.3d 296, 1 Cal. Rptr. 3d 32 (Cal. 2003).

[57] White Buffalo Ventures LLC v. University of Texas at Austin, Findlaw. Retrieved June 20, 2015.



制性要件。<sup>[58]</sup> 美国信息盗用制度现代化的典型判例摩托罗拉案的判决认为,只有被告与原告提供的产品或服务存在直接竞争,并且被告的行为减损了原告生产产品或服务的动力,以至于他的生存将受到严重威胁的时候,盗用侵权才成立。<sup>[59]</sup> 该案将“实质性”损害解释为被告利用数据的行为可能导致原告的生存受到严重威胁,颇具借鉴价值。在判断数据爬取行为是否造成“实质性”替代时,爬取的数据数量的多少,以及爬取之后对数据的利用是否有可能替代原告的产品和服务,都是应当加以考量的因素。

### 3. 爬取和利用数据的行为侵害数据主体的合法权益

消费者利益是不正当竞争判断的重要影响因素。互联网平台的数据是用户提交的或者是用户与平台互动过程中产生的,爬取数据行为如果损害用户的合法权益,如著作权或个人信息保护权,那么平台可以基于保护用户数据不受侵害的职责,阻止第三方爬取数据。但对于用户直接生成显示在网页上的公开数据,爬取行为并不一定会侵害用户的合法权益,因此平台并不能以保护用户权益为由一概禁止第三方爬取数据,只有在有证据证明爬取行为确实可能侵害用户合法权益的情况下才可以主张爬取行为不正当。

对于用户享有著作权的数据,如果第三方爬取之后在网络上传播,是需要经过著作权人授权的,因为涉及著作权人的信息网络传播权。但如果没有传播,而仅仅是抓取数据用作分析,分析的结果也没有呈现作品的原貌,是无法将其判定为侵害著作权的。对于登载在网络上供公众阅读的文章,任何人都可以阅读并以此作为新思想的来源,我们并不会因为读者的思想受到这篇文章的启发而要求读者向著作权人支付费用,机器学习也是同样。算法是机器的大脑,数据是进行运算的原材料,运算的结果如果没有呈现作品的原貌,是不会导致著作权人的权利受损的。

保护用户的个人信息免受侵害是平台的重要职责,平台经常以保护个人信息为理由拒绝第三方爬取数据,即便是公开的个人信息。但互联网用户在其公开数据中的隐私利益是不确定的,授权私人实体阻止对公开可得信息的浏览的权利会构成对公开言论和信息自由流动的巨大威胁。<sup>[60]</sup> 一个人如果选择将自己的个人资料公布在网络上,就表明他并不介意别人获取这些资料。我国《民法典》和《个人信息保护法》都规定,合理处理自行公开及其他合法公开的个人信息不承担民事责任,除非个人提出明确的拒绝或该种处理活动对个人信息重大权益存在重要影响。<sup>[61]</sup> 这说明我国对公开的个人信息并未采同意规则,而是要求数据处理者的行为合理合规。根据《个人信息保护法》规定,处理公开的个人信息之前,应当进行个人信息影响评估,确定处理行为是否会对个人权益产生重大影响,可能产生重大影响的,应取得个人同意。这样做的根本原因还是为了实现对个人信息权益保护与合理行为自由维护之间的协调与平衡。<sup>[62]</sup> 而用户在非公开个人信息中的隐私期待明显与公开信息不同。非公开个人信息中可能涉及非常敏感的个人信息,如银行卡号、密码等,应当予以更加严格的保护。因此,第三方未经授权爬取非公开个人数据,可以直接认定为不正当竞争行为。

### (三) 竞争关系考量的意义

在数据不正当竞争案件中,竞争关系往往是原、被告之间争议的焦点,因为原、被告之间

[58] [澳]马克·戴维森:《数据库的法律保护》,朱理译,北京大学出版社2007年1月版,第200页。

[59] National Basketball Association v. Motorola Inc. 105 F.3d 841(2d Cir.1997), at 853-854.

[60] hiQ Labs, Inc. v. LinkedIn Corporation, 938 F.3d 985 (2019).

[61] 参见《民法典》第1036条,《个人信息保护法》第13条。

[62] 参见程啸:《论公开的个人信息处理的法律规制》,载《中国法学》2022年第3期,第92页。

是否存在竞争关系,决定了案件是否应由反不正当竞争法调整。从我国法院对竞争关系认定的总体趋势来看,已经从狭义竞争关系过渡到广义竞争关系,这是与市场竞争和商业模式的发展变化相一致的。在市场竞争早期,经营具有替代关系的商品和服务的经营者之间会相互争夺交易机会,产生竞争关系,而如果经营的商品和服务不相同则不会存在竞争关系。但是随着企业多元化经营,尤其是网络环境下跨界竞争越来越常见,即便是非同业经营者仍会对交易机会、交易能力存在争夺,于是竞争关系的界定也向广义发展。最高法院在百度公司诉奥商网络公司案中将竞争关系界定为:如果在市场竞争中一方的行为不正当地妨碍了另一方的正当经营活动并损害其合法权益,则可以肯定二者之间存在竞争关系,<sup>[63]</sup>这表明了认定竞争关系的宽泛趋势。在当下的数据不正当竞争案中,法院多采用广义竞争关系的界定,只要数据抓取者的行为可能损害被抓取者的利益,而且同时会基于该行为获得现实或潜在的经济利益,则可以认定二者具有竞争关系。

采用广义竞争关系的概念,可以避免界定竞争关系的无穷论争,将以不正当手段获取竞争优势行为纳入反不正当竞争法规范调整。当下数据市场从产品市场进入到要素市场,数据竞争也从同类产品经营者之间提供竞品的横向竞争转化为上下游经营者之间争夺生产资料的纵向竞争,很多数据利用者抓取数据后进行衍生性使用,所提供的产品和服务与被爬取企业可能并不在一个市场进行竞争,如果囿于狭义竞争关系的理论,就很难对这些行为进行规范,而采广义竞争关系,则有利于在反不正当竞争法的框架中规范数据爬取行为。而将数据爬取行为置于市场竞争的视角之下,对其是否正当的判断纳入竞争者、消费者等多重利益考量,能够更好地平衡数据保护和公众利用之间的关系。

但是值得注意的是,在具体不正当竞争行为的认定中,某些情况下为了阻止权利过度的扩张,需要采用狭义的竞争关系,即同业竞争关系来认定。例如盗用被爬取方竞争优势的数据爬取行为,其构成要件是将爬取的数据做竞争性使用造成实质性替代的后果,这时对竞争关系应采用狭义的同业竞争关系,以避免数据权益的过度扩张。由此可见,在认定数据不正当竞争行为时,采用广义还是狭义的竞争关系要根据具体的情况,平衡好各方利益。

#### 四、数据不正当竞争规则的构建

2022年11月发布的《反不正当竞争法(修订草案征求意见稿)》第18条对涉数据的不正当竞争行为进行了规定,说明数据法律规则的构建已经提上日程。征求意见稿将保护对象界定为“商业数据”,是指“经营者依法收集、具有商业价值并采取相应技术管理措施的数据”。并对不正当获取或者使用数据的行为进行了类型化规定。从征求意见稿对“商业数据”的界定来看,似乎有意借鉴日本反不正当竞争法对限定提供数据的保护,但从不正当竞争行为的认定方式来看,征求意见稿采用的是竞争性权益保护的途径,而日本反不正当竞争法遵循的是财产权益保护的途径。以不正当获取数据为例,我国征求意见稿第18条第1款规定:“以盗窃、胁迫、欺诈、电子侵入等方式,破坏技术管理措施,不正当获取其他经营者的商业数据,不合理地增加其他经营者的运营成本、影响其他经营者的正常经营。”根据该条规定,行为人

<sup>[63]</sup> 2010年《最高人民法院公报》第8期上公布的百度公司诉奥商网络公司等不正当竞争案所确立的裁判规则:“如果二者在市场竞争中存在一定联系或者一方的行为不正当地妨碍了另一方的正当经营活动并损害其合法权益,则应肯定二者之间存在竞争关系。”

仅仅破坏技术管理措施获取他人的商业数据还不足以构成不正当竞争,还要衡量是否造成了“不合理增加其他经营者的运营成本,影响其他经营者的正常经营”的后果,这种利益衡量的认定方式是竞争性权益的保护方法。日本反不正当竞争法第2条第1款11项规定:“没有访问权限的人通过窃取、欺诈和不正当访问等违反法律规定的行为,以及同等程度的违反公序良俗的手段,破坏ID、密码和加密措施等管理手段,从所有者获取限定提供数据的行为,以及不正当获取之后使用和公开的行为,视为不正当竞争行为。”从该款规定来看,日本反不正当竞争法对破坏管理措施获取数据的行为,并没有要求后果要件,这是对财产性权益的保护方法。日本对限定提供的数据采用财产性权益保护方法,是以对“限定提供数据”进行严格界定为基础的,符合条件的数据才能作为限定提供数据进行财产性权益保护。而我国对“商业数据”虽然也做了规定,但该界定是不严格的,从该条款的上下文来看,公开的网页数据(原始数据)也是保护对象,“数据抓取协议”(Robots协议)也被认定为管理措施,这种定义的结果是无法对性质不同的数据进行区别保护。

征求意见稿18条第2、3款均以造成“实质性替代”的后果作为认定不正当竞争行为的要件,这是司法实践中“实质性替代”规则在立法中的体现。但是我国法院对“实质性替代”规则的适用,并不以数据采取了技术管理措施为前提,产生该规则的大众点评诉爱帮网案,被告并未违背Robots协议或破坏任何管理措施。实际上,对于数据爬取行为而言,如果被告爬取数据进行利用而导致实质性替代了原告的产品或服务,即无偿占有了原告的劳动成果,可认定构成不正当竞争,并不需要违反协议或者破坏技术措施的要件。

总体而言,征求意见稿第18条的问题是没有分清财产性权益和竞争性权益不同的保护方式,进而对不正当竞争行为的构成要件缺乏恰当的把握;也没有考虑不同数据的区别保护,商业数据的定义含混不清。我国要构建数据不正当竞争的法律规则,首先需要考虑的就是区分不同的数据设定不正当竞争行为的构成要件:用户直接提交或生成的展示在网页上的数据,具有原始数据的特征,应遵循竞争性权益的保护路径,平台对此类数据没有支配和控制的主观性权利,第三方爬取数据是否构成不正当竞争,要看爬取行为是否造成了实质性的损害后果,具体包括:实质性妨碍和破坏平台的正常运行、造成对被爬取方网络产品或服务的实质性替代、侵害数据主体的合法权益。而采用密码等保护措施控制并仅向特定人提供的商业化数据,则可采用财产性权益的保护路径,平台对此类数据享有支配性的主观权利,第三方未经许可获取数据,或者披露和使用未经许可获取的数据,均会构成不正当竞争行为。可以看出,司法实践中形成的“三重授权规则”是财产性权益的保护方式,而“实质性替代”和“破坏性利用”规则是竞争性权益的保护方式,在构建法律规则时,我们需要总结司法实践中的裁判经验,并对其适用情境、适用条件加以归纳和优化。

## 五、结语

数据爬取行为不正当性的认定有其复杂性,需要区分不同的情境,妥善衡量数据控制者、数据利用者、数据主体的利益。本文以司法实践中形成的裁判规则为基础,区分两种类型的数据,对其保护方式进行了分析,并在此基础上提出了认定不正当竞争行为的构成要件,以期成为当下反不正当竞争法的修订提供建议。但鉴于数据竞争行为千变万化,本文的讨论或许不能完全涵盖所有的情况,仍有待于对实践不断地探索和总结。

## Research on Rules of Law on Unfair Competition Practices for Data Crawling

Zhou Xiping

**Abstract:** What rights and interests the platform enjoys to the data is the prerequisite to judge the legitimacy of data crawling behavior. Platform data shall be protected by categories. The data generated by users directly displayed on the web page has the characteristics of originality and openness, and should be protected by means of competitive rights and interests. For commercial data controlled by the platform through password and other protective measures and provided only to certain persons, the platform may adopt the protection path of property rights. The “triple authorization rule” formed in judicial practice is the way to protect property rights and interests, while the “substantive substitution” and “destructive use” rule is the way to protect competitive rights and interests. When constructing legal rules, we should summarize the experience of adjudication in judicial practice and optimize and perfect it.

**Key Words:** Data Crawling, Unfair Competition Rights, Interests of the Data

(编辑:宋亚辉)