

数据抓取行为的反不正当竞争法规制研究

卢玉婷

四川大学法学院, 四川 成都

收稿日期: 2023年4月17日; 录用日期: 2023年6月1日; 发布日期: 2023年6月12日

摘要

数据抓取行为有利于数据共享的同时, 也成为了新型不正当竞争手段。数据抓取行为因其技术中立性和行为隐蔽性, 对其法律性质的认定应当分类讨论, 同时也增加了数据不正当抓取行为的认定难度。竞争法框架下, 可以通过互联网专条中的兜底条款、一般条款以及商业秘密保护条款对数据不正当抓取行为进行规制。由于专条兜底条款和商业秘密条款保护对象的局限性, 司法实践中常通过一般条款认定数据抓取的不正当竞争性。在司法实践形成类型规制的情形下, 有必要增设“数据抓取不正当竞争行为”条款、增强爬虫协议的法律效力。

关键词

数据抓取, 不正当竞争, 爬虫协议, 法律规制

Research on the Regulation of Anti-Unfair Competition Law for Data Capture Behavior

Yuting Lu

Law School, Sichuan University, Chengdu Sichuan

Received: Apr. 17th, 2023; accepted: Jun. 1st, 2023; published: Jun. 12th, 2023

Abstract

Data capture behavior is conducive to data sharing, but also becomes a new means of unfair competition. Because of its technical neutrality and behavior concealability, the identification of its legal nature should be classified and discussed, which also increases the difficulty of the identification of improper data grabbing behavior. Under the framework of competition law, improper data capture can be regulated through the bottom clause, general clause and trade secret protection clause in the Internet article. Due to the limitations of the protection objects of the special disclosure clause and the trade secret clause, the unfair competition of data capture is often iden-

tified through the general clause in judicial practice. In the case of the formation of type regulation in judicial practice, it is necessary to add “data grabbing unfair competition behavior” clause and enhance the legal validity of crawler agreement.

Keywords

Crawl Data, Malfeasant Competition, Robots Exclusion Protocol, Legal Regulation

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在数字经济时代，数据的商业价值日益凸显，已然成为互联网企业的核心竞争力所在。谁能掌握数据，谁就掌握了财富密码。经营者力图通过各种手段获取数据信息，以期攫取经济利益、占据竞争优势。由于数据抓取的高效性、低廉性，数据抓取已经成为部分企业收集数据的首要手段。然而，数据不正当抓取情况愈演愈烈，已经成为不正当竞争新型化趋势。国内司法实践中，数据抓取导致纠纷频出，争讼不止。目前，我国《反不正当竞争法》尚未对数据抓取行为进行明确规定，司法裁判的认定路径和标准不一，削弱了法的可预测性，对于数据抓取行为无法有效遏制。为此，本文立足反不正当竞争法的理论基础，检视现有数据抓取的竞争法规制路径，并提出完善建议。

2. 数据抓取行为概述

2.1. 数据抓取行为相关概念界定

数据抓取行为是指，利用“网络爬虫”技术，有目的地、自动地、精准地搜索并采集网页、文档、图片等具有商业价值的数据内容，并进行筛选和归纳的行为。网络爬虫(Web Crawler/Spider)，是数据抓取行为的技术核心，其实质上是一种计算机程序或脚本，能够反复执行设定的程序，模拟网络用户的浏览行为进行信息调取和信息处理，从而达到收集数据的目的。

2.2. 数据抓取行为的特征

2.2.1. 技术中立性

数据抓取行为其具有技术性，网络爬虫的合理使用不为法律所禁止。在信息爆炸的时代，网络爬虫技术的应用有利于高效便捷、有针对性地实现数据信息的获取。同时，被抓取数据方的数据也因数据曝光度增加而获得流量，增加其市场竞争力，实现双方的共赢。网络爬虫技术在诸多领域已经实现良性发展，节省了寻找和搜集数据的时间、精力和金钱。网络爬虫的应用领域已经从搜索引擎，逐步扩展到电商、社交平台，覆盖我们生活的全方面，有利于社会经济的发展和数据共享。然而，技术的中立并未能实现数据抓取行为的完全正当性[1]。数据抓取行为导致被抓取数据网站访问量过大而导致崩溃、企业核心商业数据被恶意抓取的情形时有发生，为了防止数据抓取的不利后果以及维护数据利益，限制数据抓取技术应运而生。《互联网搜索引擎服务自律公约》为规制网络爬虫行为设置了“Robots”协议，以此明确数据的可爬取性，爬虫自动排除网站不愿意被抓取的内容，以期维持数据抓取行为的合法性和正当性。

2.2.2. 行为隐蔽性

数据抓取行为外观上与用户浏览活动一致,数据抓取方通过少量、多次抓取数据,其行为在分秒之间即可完成,并且可以在后续进行技术处理为其数据抓取行迹披上隐蔽外衣。因此,互联网竞争模式下,数据抓取方和数据被抓取方的攻防存在“敌明我暗”的不平衡性。从技术角度而言,防守方只能通过技术降低爬虫爬取数据的效率,不能从实质上防止数据爬取。而且,数据抓取行为的实质目的往往通过前端数据收集行为掩盖后端不正当竞争目的。表面上数据抓取行为只进行了数据收集,但是实际上是为了实现诸如篡改用户浏览器主页、对不同厂商浏览器进行区别对待、算法歧视、数据“搭便车”等抓取效果[2]。

2.3. 数据抓取行为的法律性质

由于数据的商业价值和竞争性,数据抓取方和数据被抓取方二者存在利益对立,既要保护数据权益,又要实现数据共享,防止垄断行为的发生。合法、合理的数据抓取行为有利于实现数字共享,促进我国数字经济现代化发展。然而,恶意网络抓取行为超越了技术中立性,抓取其他经营者的核心竞争数据并攫取其竞争优势,触碰了法律的红线。由于“Robots”协议作为不具有强制性的互联网行业惯例,其未能阻止数据恶意抓取行为。随着技术的进步,网络爬虫技术可以通过数据解密、验证码自动识别等技术突破反爬技术手段,从而获得禁止爬取的数据内容。这种以破坏他人的竞争优势,转化为自己的竞争优势,以低廉成本攫取他人经营成果的“搭便车”行为,存在构成不正当竞争的法律风险。

3. 数据抓取行为构成反不正当竞争的现有认定路径

3.1. 适用互联网专条

针对互联网新型业态和经营模式,我国《反不正当竞争法》第12条对利用网络技术手段对其他经营者正常提供产品和服务的行为造成妨碍、破坏的行为进行了规制。互联网时代竞争形式复杂多变,为了克服法律的滞后性,该条款采用列举加兜底的立法方式。列举条款对互联网领域的类型化不正当竞争行为进行立法确认,但并未对数据前沿问题予以回应。

从数据抓取的行为特征和使用领域来看,其应当属于互联网专条兜底条款的规制范围。然而,该兜底条款警示意义大于实践意义,难以对数据抓取行为进行规制。一方面,互联网专条中的兜底条款出台后,其尚未形成权威性的认定构成要件,实践中适用较为困难。法官说理进行法律判断的依据不足,主观性较强,难以得到纠纷双方的一致认可。另一方面,数据抓取行为具有技术中立性,造成“妨碍”“破坏”的认定难以精准解决当下数据不正当抓取行为。数字经济下,数据共享已经常态化,数据抓取技术的进步,导致其并不会影响其他经营者正常经营。通过造成“妨碍”“破坏”认定构成不正当竞争,存在适用局限性。

3.2. 适用反不正当竞争法一般条款

我国司法实践中,大多引用《反不正当竞争法》第2条认定数据抓取行为构成不正当竞争。关于该条款已经形成了完整权威的认定构成要件,在认定数据不正当抓取中得到普遍适用。由于数据抓取行为的特殊性,在认定过程中主要存在以下问题:

3.2.1. 经营主体的认定

不正当竞争的主体只能是经营者。司法实践中根据是否有营业资质认定经营者主体身份的做法存在一定偏颇。通过行政许可的手段能够减少经营者认定的程序,但是缩小了经营者的范围,不利于市场竞

争秩序的维护。经营者主体身份的认定应当从行为出发,只要从事了经营行为,则具备经营者身份。不应以其是否具备经营者资格而予以认定,非法经营的经营者也是经营者。同时,在互联网领域,经营者的提供的产品或服务并不具有营利性时,是否影响对其经营者身份的认定?在“酷米客”诉“车来了”案件中,二者均提供免费提供公交实时查询服务的服务,用户可以免费下载和使用软件,即经营者提供的服务不具有营利性。“车来了”APP利用网络爬虫技术大量获取并无偿使用“酷米客”APP实时公交信息数据,被认定为构成不正当竞争行为。法院认为,提供某项产品和服务的营利性不是认定经营者主体身份的标准,二者均为不正当竞争法意义上的经营者[3]。

3.2.2. 竞争关系的认定

竞争关系通常被认为是构成不正当竞争的前提。大型互联网公司的商业版图往往涉及多个领域,同业经营的边界模糊。在大众点评诉百度案中,百度地图主要提供定位、导航等服务,但还提供商户信息和评价;大众点评作为一个消费点评网站主要提供商户信息、用户评价等,二者在提供商户信息和点评信息上有交叉部分,存在竞争关系[4]。

发生在非同行竞争者之间的数据抓取行为是否能认定为不正当竞争?在新浪微博诉脉脉案中,微梦公司运营的新浪微博属于互联网社交媒体平台,而淘友公司运营的“脉脉”属于实名制商业职场社交应用,二者属于非同行竞争者。但是二者使用同样的商业模式,即吸引用户使用平台或软件,使用过程中产生用户信息和用户使用偏好等信息,此类信息从整体上构成商业数据,通过安全地使用涵盖用户偏好的商业数据给互联网企业增加用户粘度,便于经营者在此基础上开展多种业务获利。新浪微博平台用户数量庞大,储存大量用户个人信息与浏览记录等信息,由此可以得出用户的职业背景、兴趣偏好等商业数据,这些数据对于想要实现商业社交平台的脉脉来说,具有很强的经济价值。法院认为,双方在对相关用户社交类信息的使用等方面存在竞争利益,具有竞争关系。由于互联网行业跨领域交融的特性以及获利模式,竞争关系已经发生实质性扩张变化,应从广义的竞争关系进行认定,而不应限于同业经营者之间。

3.2.3. 行为的不正当性

行为的不正当性通常体现在违反了诚实信用和商业道德。从诚实信用角度出发,主要考量利益。经济法的诚实信用原则实际上来源于民法,其主要是指经营者以善意的方式进行竞争,不损害他人竞争利益来谋取自己的利益。在新浪微博诉脉脉案中,法院以数据被抓取方新浪微博获得和维护数据的成本巨大,认定数据抓取方脉脉“搭便车”行为不正当性。同时,二审法院在论述淘友公司违反诚实信用原则和商业道德时指出:“互联网中,对用户个人信息的采集和利用必须以取得用户的同意为前提,这是互联网企业在利用用户信息时应当遵守的一般商业道德”,并提出对于用户类商业数据的采集和利用应当遵守“用户授权”+“平台授权”+“用户再授权”的三重授权原则[5]。

从商业道德角度出发,实务中对于数据抓取行为不正当竞争的认定,主要考量是否遵守行业惯例和数据协议。“robots协议”作为行业自律文件往往成为并不具备法律效力。“robots协议”和数据协议本身也可能存在不正当性,增加了数据抓取行为不正当性认定的难度。值得注意的是,robots协议并非认定不正当竞争的必要条件。在大众点评诉百度案中,大众点评未设置robots协议限制抓取用户点评信息,百度抓取用户信息的行为未违反大众点评的robots规则。但法院认为,百度公司大量、全文使用汉涛公司耗费大量成本运营且具有核心竞争力的涉案点评信息,违反了公认的商业道德和诚实信用原则,具有不正当性。百度进行数据抓取的违法行为并未以没有违反robots协议而发生实质改变。由此可见,robots协议并非认定不正当竞争的必要条件,需具体考量其数据抓取行为以及抓取之后的使用行为的法益侵害性,全面地认定数据抓取行为是否构成不正当竞争。

3.2.4. 造成实质损害

大众点评诉百度案中，法院认为，“百度公司的行为给汉涛公司造成了实质损害，具有不正当性，构成不正当竞争”。在新浪诉脉脉案中，法院也以淘友公司对公平竞争秩序和微梦公司竞争利益的损害为条件认定行为的不正当性。实际损害主要包括对市场竞争、经营者和消费者利益的损害。同时，还要求数据抓取行为与造成实际损害的结果之间具有因果关系。对于市场竞争秩序和竞争者利益的损害，主要体现在抓取他人具有核心竞争力的数据，通过诸如提供可替代产品或服务的行为，损害他人的竞争优势，扩大自己的竞争优势。

在新浪诉脉脉案中，法院也考虑到了用户数据未经允许被抓取时，消费者利益受到的损害。数据抓取行为对消费者利益的损害，主要是在于消费者个人信息的泄露、滥用。一般而言，在数据被抓取方获得的用户数据需经用户授权。由于数据抓取行为的隐蔽性，在数据被抓取方和用户未能察觉的情况下，抓取消费者诸如个人基本资料、使用偏好的信息，对消费者的个人信息保护权利造成了实质损害。

3.3. 适用商业秘密条款

数据抓取行为适用我国《反不正当竞争法》第9条，需要符合以下构成要件：

3.3.1. 被抓取的数据是商业秘密

被抓取数据具有秘密性。秘密性要求被抓取的数据应该处于不为公众所知悉的保密状态，且一般人不易通过正当途径获得或方法获得。从主观上来讲，数据持有者具有对该数据予以保密的主观愿望。而判定是否存在主观秘密性，主要是看数据持有者是否采取了相应的保密措施。例如，在对涉密数据设置密码、签订保密协议。从客观上来讲，数据没有被公众所了解以及不能从公开渠道直接获取。这时需要对该数据是否为公众所了解的事实以及信息获取的难易程度进行考察。例如，“酷米客”诉“车来了”案中，汉光公司所爬取的“酷米客”的实时公交信息数据并非未公开的，它可免费提供公众查询，显然不构成不为公众所知悉，不是商业秘密。因此，公开的数据不符合商业秘密的特征，其被抓取不能够适用商业秘密保护条款。商业秘密保护条款不能完全覆盖数据抓取行为的不正当情形。

被抓取数据具有非物质性和商业价值性。数据作为一种无形资产，其可以通过载体表现出来，但是其内涵的思想内容却是无形的。也就是说，数据的商业价值不在于载体，而在于信息。在互联网时代，数据日益成为重要的商业资源，能够作营利性使用，并给持有人带来经济利益及竞争优势。例如，对于用户偏好数据的掌握，往往成为企业进行创新和数据开发的重要风向标，研发在市场上获得受众的喜爱产品。由此可见，数据蕴含的信息具有极大的商业价值，能够使得企业在市场竞争中脱颖而出，占据利好。

3.3.2. 数据抓取行为属于侵犯商业秘密行为

我国《反不正当竞争法》规定“电子侵入”属于侵犯商业秘密的行为。从行为外观上看，数据抓取利用爬虫技术属于电子数字手段。而且，“侵入”手段具有不正当性，要求数据抓取行为是在违反“Robots”协议、违反保密协议或破坏他人保密技术等未经他人许可的情形下获取的。

4. 数据抓取行为的反不正当竞争法规制路径的优化

4.1. 增设“数据抓取不正当竞争行为”条款

在互联网专条兜底条款意义性大于实践性、一般条款使用需兼顾数据抓取个性、商业秘密保护条款针对对象有限的情形下，笔者认为有必要增设“数据抓取不正当竞争行为”条款。数据日益成为重要竞争资源，数据抓取构成不正当竞争案例频出，其在司法实践已经成为类型化案例。增设“数据抓取不正

当竞争行为”条款，有利于减少原则性一般条款的使用，增强法的确定性，构建具体的行为准则，维护互联网平台竞争的有序性。

4.2. 增强“Robots”协议的法效性

“Robots”协议作为互联网行业自律规范，其并不具有法律效力。确认“Robots”协议作为认定商业道德标准的法律地位，有利于弥补现行法律的不足、增强司法实践中法官释法的依据、促进行业自律。同时，还应规范“Robots”协议的适用。如前文所述，滥用“Robots”协议来规避正常的数据抓取行为，无疑会对数据抓取技术创立初衷、其他经营者利益和市场竞争秩序造成损害。

5. 结语

在数据具有极高商业价值的互联网经济背景下，数据抓取行为应运而生，因其技术中立性，将其扬长避短有利于我国数字经济的发展和创新。然而，数据抓取行为的违法性提高，使其逐步脱离了中立的技术定义，背离了科技伦理而发生了“异变”，成为新型不正当竞争行为。现有数据抓取行为构成反不正当竞争的现有认定路径仍然存在一定的缺陷，“Robots”协议规制有限性下的行业自律问题尚未得到解决。亟需弥补完善相应法律法规，增强“Robots”协议的法效性和规范性，引导数字经济正向发展，为数据竞争的规制问题探寻本土化的解决方案。

参考文献

- [1] 陈兵. 保护与竞争: 治理数据爬取行为的竞争法功能实现[J]. 政法论坛, 2021, 39(6): 18-28.
- [2] 蔡川子. 数据抓取行为的竞争法规制[J]. 比较法研究, 2021(4): 174-186.
- [3] 杨梦婷. 爬虫抓取数据行为的反不正当竞争法规制研究[D]: [硕士学位论文]. 重庆: 西南政法大学, 2020.
- [4] 张奉祥. 论商业数据抓取行为的竞争法规制[D]: [硕士学位论文]. 北京: 中国政法大学, 2020.
- [5] 徐菁. 数据抓取行为的竞争法规制[D]: [硕士学位论文]. 湘潭: 湘潭大学, 2020.