

论互联网数据抓取行为的反不正当竞争法规制

黄金金

浙江理工大学法政学院, 浙江 杭州

收稿日期: 2023年7月4日; 录用日期: 2023年7月21日; 发布日期: 2023年9月12日

摘要

随着时代的发展, 数据抓取行为已经成为企业获取数据的一个重要的方式, 然而, 数据抓取行为存在的法律风险往往被忽视。新浪微博诉脉脉案被称为大数据引发的不正当竞争第一案, 对于互联网数据抓取行为的反不正当竞争法规制具有重要的参考意义。在司法实践中, 针对数据抓取行为, 《反不正当竞争法》的适用存在诸多困境, 商业秘密条款往往被司法人员忽略, 网络条款难以涵盖所有的数据抓取行为范式, 一般条款被过度的引用。对此, 我们需要明确数据抓取行为不正当性的判断标准, 明确具体条款的选择适用规则, 明确商业秘密条款与网络条款的适用情形。

关键词

数据抓取, 不正当竞争, 法律规制

On the Anti-Unfair Competition Laws and Regulations of Internet Data Scraping

Jinjin Huang

School of Law and Politics, Zhejiang Sci-Tech University, Hangzhou Zhejiang

Received: Jul. 4th, 2023; accepted: Jul. 21st, 2023; published: Sep. 12th, 2023

Abstract

With the development of the times, data scraping behavior has become an important way for enterprises to obtain data; however, the legal risks of data scraping behavior are often overlooked. The Sina Weibo v. Pulse case is known as the first case of unfair competition caused by big data, and has important reference significance for the anti-unfair competition laws and regulations of Internet data scraping. In judicial practice, there are many difficulties in the application of the Anti-Unfair Competition Law to data scraping, trade secret clauses are often ignored by judicial per-

sonnel, online clauses are difficult to cover all paradigms of data scraping behavior, and general clauses are overly quoted. In this regard, we need to clarify the criteria for judging the impropriety of data scraping, clarify the rules for selecting and applying specific clauses, and clarify the application of trade secret clauses and network clauses.

Keywords

Data Scraping, Unfair Competition, 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. 问题的提出

随着时代的发展,科学技术的飞速提升,大数据时代已悄然来临。互联网领域中,数据的价值变得愈发重要,越来越多的商家意识到,抓住数据就意味着抓住了竞争的优势,就能在“商业战争”中占据有利的地位。在这种背景下,如何获得更多更优质的数据就成为了商家们所追求的目标,对于数据的竞争战也开始悄悄打响。数据抓取就是获得数据的一种重要方式,围绕数据抓取的竞争行为也逐渐成为互联网时代的一种全新的竞争方式。互联网背景下,数据抓取即是指对数据的收集,经营者通过各种方法从用户或者其他经营者处收集到数据来对自身的运营产生积极的影响。其中经营者对其他经营者收集到的数据进行抓取就涉及到了不正当竞争的问题,针对此种行为,就需要依靠反不正当竞争法来对经营者的权利进行保护。

新浪微博诉脉脉案被称为大数据引发的不正当竞争第一案,对于互联网数据抓取行为的反不正当竞争法规制具有重要的参考意义。新浪微博诉脉脉案¹中,原告新浪微博与被告脉脉软件都提供给用户社交服务,同时,二公司曾签订协议,约定以 OpenAPI 的方式由脉脉向新浪微博处抓取数据,并且,双方签订了《开发者协议》。但在该协议签订的内容和时间之外,脉脉对新浪微博的用户数据进行了收集并展现在脉脉中,为此,新浪微博向法院提起诉讼,主张脉脉软件非法抓取、使用新浪微博用户数据。一审法院认为,脉脉的行为不符合《开发者协议》的约定,该数据抓取行为既没有合同上的约定,也缺乏法律上的依据,并且未受到来自用户的同意,此种情况下,脉脉抓取微博用户的数据信息危害了新浪微博的数据安全,构成不正当竞争。在二审的判决书中,二审法院引用了大量《开发者协议》的内容以及众多专家学者的意见,在关于不正当竞争行为的认定中,法院引用的法条为《反不正当竞争法》的第二条,即反不正当竞争法中的“一般条款”,认为脉脉的行为违反了诚信原则以及公认的商业道德,故构成了不正当竞争。无论是一审法院还是二审法院,都高度强调了数据保护的重要性。新浪微博诉脉脉案二审的最终判决于 2016 年的年末做出,此后《反不正当竞争法》经历了两次修改,反不正当竞争法司法解释也于今年三月份正式施行,在当下时代,最新的《反不正当竞争法》中的条款对该案以及其他类型的数据抓取行为是否能够进行规制?解决了这个问题后,我们方能对互联网数据抓取行为的反不正当竞争法规制描述出一个大致的轮廓。

¹北京淘友天下技术有限公司、北京淘友天下科技发展有限公司与北京微梦创科网络技术有限公司不正当竞争纠纷案,原告北京微梦创科网络技术有限公司是“新浪微博”网站及软件的运营公司(简称“新浪微博”);北京淘友天下技术有限公司、北京淘友天下科技发展有限公司是“脉脉”网站及对应软件的运营公司(简称“脉脉”)。参见北京淘友天下技术有限公司、北京淘友天下科技发展有限公司与北京微梦创科网络技术有限公司不正当竞争纠纷上诉案(2016)京 73 民终 588 号。

2. 互联网数据抓取行为概述

2.1. 数据抓取行为的概念

通常认为,数据的生命周期包括数据收集、数据整合、数据分析以及数据利用四个阶段,数据抓取亦或是数据收集正是数据生命周期的第一阶段[1]。根据收集对象来源的不同可以分为两类:第一类是从用户(不限于人,也包括金融、经济等数据的数据主体)处收集原始数据,其面对的是海量的主体,并且需要把信息转换成编码形式显示的数据。比如淘宝收集用户对商品浏览、收藏、交易的痕迹,网站收集的特定地区包括房价等指标在不同时期变化情况的房地产数据集等。第二类是收集其他经营者收集自用户处的数据。该类数据既可以是他人整理并发布的数据集,也可以是未经处理的零散数据内容。新浪微博诉脉脉案中,脉脉和新浪微博都属于社交应用,双方曾签订《开发者协议》,通过微博平台 OpenAPI 进行数据交易,其具体措施主要有两项,第一是抓取使用了新浪微博平台的信息,这个信息主要是指用户的交友信息、职业信息等公开的一些信息,第二是脉脉的用户手机通讯录的联系人上传之后,脉脉利用抓取电话号码的方式,进行和微博用户的匹配,实现对应关系。脉脉的数据抓取行为则是属于第二类通过其他经营者收集用户数据的行为,实际上,涉及不正当竞争的互联网数据抓取行为基本上属于此类。

2.2. 数据抓取行为的方式

目前主流且合法的网络数据收集方法有三种:开放数据集下载、API 读取、爬虫。

开放平台数据集下载是第一种网络数据收集方式。该种方式通常包括收费数据和免费开放数据两种数据类型,在平台注册之后即可以下载免费开放数据,根据目标平台的不同所下载的数据包会有 CSV、Excel、JSON、XML 等不同的格式,下载完成后可以使用 Python 工具来提取和整理数据,使数据以表格或者图表的可视化形式呈现。

利用 API (Application Programming Interface)读取数据是第二种网络数据收集方式。是由目标网站主动提供一个通道,当他人需要某一部分数据的时候,虽然没有已经整理完成的数据集,却可以利用这个通道描述自己想要的数据库,然后网站审核(一般是自动化、瞬间完成的)之后,认为可以提供的,就会立刻把先前明确索要的数据发送过来。新浪微博诉脉脉案中,脉脉即是通过 API 的方式抓取了新浪的用户数据。开放平台经营者通过 Open API 的方式向第三方提供数据和服务时,其为了保证系统和数据安全,有权通过开发者协议、开放平台规则等规范选择数据开放范围以及第三方应用获取数据的使用方式、使用范围等。授权决定了数据抓取行为的合法性边界,也因此数据抓取者是否获得了足够的授权就成为了最重要的考量因素。

利用网络爬虫爬取数据是第三种网络数据收集方式。网络蜘蛛是一个很形象的名字。把互联网比喻成一个蜘蛛网,那么 Spider 就是在网上爬来爬去的蜘蛛。网络蜘蛛是通过网页的链接地址来寻找网页。从网站某一个页面(通常是首页)开始,读取网页的内容,找到在网页中的其它链接地址,然后通过这些链接地址寻找下一个网页,这样一直循环下去,直到把这个网站所有的网页都抓取完为止。如果把整个互联网当成一个网站,那么网络蜘蛛就可以用这个原理把互联网上所有的网页都抓取下来。从这方面看来,网络蜘蛛就是一个爬行程序,一个抓取网页的程序。

3. 《反不正当竞争法》适用的法律困境

很明显,并非所有的数据抓取行为都是不正当竞争行为,新浪微博诉脉脉案中,法院就不断的强调新浪微博作为市场的优势方应承担起相应的责任,而新浪微博在数据保护上出现了差错,在归责方面法院就可能会倾斜于被告方。反不正当竞争法设立的目的即是促进社会主义市场经济健康发展,保护经

营者和消费者的合法权利，因此，合理、合法的数据抓取行为属于市场调节下的一种功能，法律不应去过多干涉，而对于市场不公平的竞争，不合理的数据抓取行为就需要用到《反不正当竞争法》，来判断其是否合法，该承担怎样的责任。可以预见，随着时代的发展，互联网数据抓取行为将会变得普遍，此类案件也会逐渐增多。在面对互联网数据抓取案件时，我们可以根据《反不正当竞争法》的具体规定如商业秘密条款、网络条款和一般规定如诚实信用原则、商业道德以及以往司法裁判的案例来进行规制，但在目前的司法实践中，仍然存在不少问题，可以看出，互联网数据抓取不正当竞争行为的法律规制存在如下困境。

3.1. 停留于纸面的商业秘密条款

在新浪微博诉脉脉案中，法院对于数据抓取行为仅从《反不正当竞争法》的第二条予以分析，对商业秘密条款避而不谈，可能一部分原因就在于技术信息与经营信息较难与用户的交友信息、职业信息等产生联系，或者是较难界定，但经过修订后，商业秘密的解释得到了扩大，商业秘密不再局限于技术信息和经营信息。新浪微博收集到的用户信息虽然单看个别可能会得出司法解释中“可以从其他公开渠道获得”的结论，从而认定出其不构成商业秘密，但我们不能忽略这些用户信息是作为一个整体的数据被脉脉所抓取，从整体上来看，显然这些众多的用户信息是几乎无法完整的从其他渠道获取，并且是需要付出大的代价来获取的，应属于不为公众所知悉。新浪微博通过收集这些用户信息并对用户进行正反馈，提高用户的粘性，从而对平台产生积极的影响，而且新浪微博与脉脉签订的《开发者协议》本身就证明了该数据具有商业价值。新浪微博通过 OpenAPI 的方式与脉脉进行数据交易，虽然二审法院指出其保密措施有所欠缺，不够到位，但这并不能忽略其采取了相应的保密措施这一事实。综上，我们或许可以从商业秘密这一条款中获得新浪微博诉脉脉案的另一种裁决思路。商业秘密条款也是判断数据抓取不正当竞争行为的有力工具，却往往被司法人员所忽略，新法对于商业秘密条款作出了一系列的重大修改，如何将这些修改后的商业秘密条款与目前的司法实践相结合也是目前法律规制的一大困境。

3.2. 无法涵盖数据抓取行为范式的网络条款

《反不正当竞争法》在新修订后，新增了第十二条“网络条款”，这一条款选择了与其他条款不同的立法模式，即“概括 + 列举 + 兜底”。首先，“概括性条款”明确了互联网经营行为应当受到整个《反不正当竞争法》的规制，而不仅仅是网络条款的规制；其次，“列举性条款”明确列举了三种互联网的不正当竞争行为^[2]；“兜底性条款”则是在最后将其他可能的不正当竞争行为进行了概括。目前主流的网络数据收集方法有三种，即开放数据集下载、API 读取、爬虫。这三种已经难以界定到三种类型当中，而随着网络技术的不断发展，会出现越来越多的数据抓取方式，直接抓取网页源码、模拟浏览器操作、或者利用数据抓取的专用工具软件等等，这些或符合三种类型化行为，可以直接进行判定，或许不能，这时就可能用到兜底性条款，如果兜底条款也不适用，那最后才会用到一般条款。立法者设立“兜底性条款”对“列举性条款”进行补充，期望能够通过赋予司法机关一定自由裁量权，在面对互联网竞争行为时，不至于陷入无法可依的局面，也在一定程度上摆脱了对“一般条款”的依赖。然而，现代社会的瞬息万变，层出不穷的数据抓取方式，越来越难以界定到“列举性条款”当中，对“兜底性条款”的设立可能将使得行政机关和司法机关在界定不正当竞争行为时拥有较大自由地带，种种原因造成“兜底条款”逐渐变为“万能条款”的无奈局面^[3]。

3.3. 具体条款与一般条款的竞合问题

当今市场竞争程度日益激烈并趋向复杂化、多样化，《反不正当竞争法》的修订虽然对市场行为的

规制已经起到了很大的积极作用，但是现有的法律条文对于网络环境下的不正当竞争仍然有许多不适应的地方，而数据抓取又是在网络环境中也难以处理行为。现有的数据抓取案例中，多采用的是《反不正当竞争法》中的一般条款，极少运用的是具体条款。但是如果把法律比作一栋大楼，那么一般条款应该起到的是地基的作用，频繁的去使用它并不利于法治的发展，适用一般条款应是具体条款无法适用时而“不得不去”使用，法律的要求应是具体的。

在具体条款与一般条款的竞合问题中，尤以网络条款和一般条款的竞合关系最为突出。在《反不正当竞争法》修订之前，解决互联网竞争行为纠纷主要依赖于一般条款的规制，法官的自由裁量权较高，而新法修订后，司法人员裁判案件则会面临一般条款与网络条款的选择适用问题，当出现不正当竞争行为时，在触犯一般条款的同时，还可能会触犯到网络条款，网络条款的适用自当摆在一个更优先的位置，然而这种竞合问题本身就往往被忽略，因此，一般条款与网络条款的适用条件以及适用路径也是目前司法实践亟待解决的问题[4]。

不管是具体条款还是一般条款，都需要正确的去适用，相对而言，一般条款的适用更加依赖于法官的法律素质，在实务中出现的问题也较多。具体条款需要与时俱进、尽可能的覆盖到日益增加的不正当竞争行为中去，不断修订的《反不正当竞争法》也是朝着这个目标前进。

4. 互联网数据抓取行为的法律规制建议

4.1. 明确商业秘密条款的适用条件

在司法实践中，对于被诉行为是否属于侵犯商业秘密行为的判断，一般遵循以下思路：第一，权利人主张的信息是否属于其商业秘密；第二，被告是否属于《反不正当竞争法》第九条调整的对象；第三，被告所使用的信息是否为权利人的商业秘密；第四，被诉行为是否属于《反不正当竞争法》第九条规定的行为。新的《反不正当竞争法》第九条第一款第(一)项的规定中新增经营者不得实施以“电子侵入”的方式获取商业秘密，这是为顺应现行数字经济环境下依托于新技术衍生出的侵权行为类型，而这一款对于数据抓取不正当竞争行为的认定具有重大意义。此款规定意味着只要证明经营者所抓取的数据属于商业秘密，并且其使用的手段属于“电子侵入”或者其他不正当手段，那么就能够用《不正当竞争法》来对此种行为进行规制。关于“电子侵入”，《反不正当竞争法》及其司法解释未予以明确定义，但从司法实践以及学者的意见来看，侵入的客体应当做广义的理解，只要是与互联网有关的载体都可以被纳入到该客体当中，例如云盘、网站、数据库等[5]。关于商业秘密，不为公众知悉、具有商业价值、经过权利人采取了相应的保密措施以及属于商业信息，这是新的《反不正当竞争法》对商业秘密作出的详细定义，已不再局限于技术和经营信息的商业秘密显然对于数据抓取的规制更相适应。当满足这些条件时，就可以在司法实践中尝试利用商业秘密条款来对数据抓取行为予以规制。

4.2. 将数据抓取行为纳入“网络条款”的规制范围

网络条款明确列举的三种不正当竞争的行为对于判断数据抓取的合法与否也有着重要的作用。以网络爬虫的方式为例，当数据抓取者采用植入链接等方式对目标网页进行干涉时，就有可能触犯到网络条款所具体列举的第一项规定。当以数据抓取的方式来掩盖自己的非法目的时，如假借抓取数据为名，对用户进行误导、欺骗，使其关闭、卸载数据拥有者的网络产品或服务，则当然构成了不正当竞争。网络条款的规制路径可能更加复杂，但用好这柄利器，对于数据抓取行为的法律规制将大有裨益。

然而，如前所述，现有的“网络条款”内容已无法有效的覆盖新型不正当竞争的主要类型。考虑到数据抓取行为与互联网经营行为有所交叉，我们可以明确数据抓取行为的构成要件，在“网络条款”下新增一个“数据抓取条款”或者“数据条款”，将数据抓取行为纳入“网络条款”的规制范围，这同样

可以契合“网络条款”的立法模式。例如，首先明确数据抓取行为应当受到《反不正当竞争法》的全文规制。其次通过对数据抓取行为的分类，明确列举三种数据抓取不正当行为，例如：未经用户、企业的授权通过 OpenAPI 抓取其他企业的用户数据；通过网络爬虫抓取数据违反相关协议；通过数据集下载的方式抓取数据而未支付对价等描述。最后设置兜底条款，其他妨害、破坏其他经营者经营的数据抓取行为。如此便可最大程度的减少司法实务人员对“兜底条款”的依赖。对于“兜底条款”的适用也应格外谨慎，只有充分的证据能证明其数据抓取行为确实妨碍了他人的经营活动，破坏了正常的市场经济秩序等才能被认定为是不正当竞争行为。法院在处理案件时应当严格限制第十二条“兜底性条款”的适用，避免“一兜到底”现象，鼓励市场良性竞争与严惩网络不正当竞争行为之间的关系应当被正确处理，防止造成对市场创新能力与创新热情的破坏。当然考虑到法律的稳定性，此种规定亦可规定在相关的司法解释中，这样我们或许就可以将“三重授权”原则、Robots 等“行业惯例”的问题纳入到法律条款当中，这些司法解释也都可以作为法院裁判的依据[6]。

4.3. 谨慎使用一般条款

《反不正当竞争法》第二条作为一般条款不可避免地具有双刃性，其赋予的自由裁量权难免会带来相应的弊端，因此，如何保证一般条款的合理适用，如何保证一般条款带来的开放性与扩张性二者的平衡，是一般条款在司法实践中面临的重大问题。根据一般法理，一般条款的适用应得到严格的限制，只有当穷尽所有具体条款而仍不能对行为予以规制时才能够用到一般条款。在适用一般条款时，应当以公认的商业道德和诚信原则为核心，排除其他无关因素的干扰，以此来防止司法权的滥用[7]。对一般条款的适用应当严格控制，并加以一定限制，防止其产生过度滥用问题。

援引“一般条款”作为判决依据应当是在立法不完备，无法可依的前提下。适用一般条款的情形不应成为司法常态，必须对其适用前提进行严格审查[8]。德国立法在这一问题上的经验值得借鉴，德国《反不正当竞争法》严格规定了一般条款的适用条件，法律所标注的具体条件没有全部符合的情况下，禁止一般条款的适用。并且对于可以解释一般条款进行判决的机构和方式也作出了严格的规定，即联邦最高法院通过判决的刑事对一般条款的适用予以解释，以此保障法律的稳定性与权威性。

我国各地市场、经济、司法环境差异较大，各地区以及各层级的法官的法律素养亦有着较大的差别。这些都可能造成“同案不同判”的结果，导致司法不公，损害了法律的权威，在现实中已经有相关的对抗型案例存在。另外，低层级的法院还容易受到当地政府地方保护主义的影响[9]。因此，对于《反不正当竞争法》第二条的司法适用应该愈加谨慎，将法条具体化、释法等的权利交由最高人民法院行使，而非由地方各级人民法院随意的“以案释法”。

5. 结论

竞争无处不在，在互联网经济迅猛发展的今天，竞争的战场已经由传统领域逐渐延伸到虚拟经济市场中，互联网环境下经营者经营方式更加灵活、经营领域更加广泛，也为隐藏在技术手段之下的不正当竞争提供了更多的可能。互联网经济环境下，数据之争越来越成为商家必争之地，近年来一系列用户数据竞争案件的频发，说明对用户数据生态竞争环境的规制已刻不容缓。反观我国现行的法律规制体系，“具体条款”的适用仍有较大的局限性，数据抓取行为的不正当性判断标准仍较为模糊。

针对上述问题，本文在论述相关法律、法规的规制难点之后，借鉴外国相关立法经验，结合我国司法实践，立足于《反不正当竞争法》现行立法模式寻找突破点，首先建议增加数据抓取不正当竞争的列举条款，明确不正当竞争行为的判断标准；其次，司法人员也应谨慎使用一般条款与兜底条款；最后，应当合理的选择适用商业秘密条款与网络条款。

参考文献

- [1] 韩伟. 数字市场竞争政策研究[M]. 北京: 法律出版社, 2017: 154.
- [2] 李兆阳. 《反不正当竞争法》视角下对数据抓取行为规制的反思与修正[J]. 暨南学报(哲学社会科学版), 2021, 43(6): 65-76.
- [3] 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.
- [4] 沈贵明, 刘源. 数据抓取行为反不正当竞争法规制困境与对策[J]. 中国流通经济, 2021, 35(1): 89-96.
- [5] 汪东升. 论侵犯商业秘密罪的立法扩张与限缩解释[J]. 知识产权, 2021(9): 41-55.
- [6] 郑友德, 张钦坤, 李薇薇, 等. 对《反不正当竞争法(修订草案送审稿)》的修改建议[J]. 知识产权, 2016(6): 3-22.
- [7] 赵丹, 沈澄. 数据抓取不正当竞争纠纷的司法审查要素考察与反思[J]. 科技与法律(中英文), 2023(2): 52-59.
- [8] 李慧敏, 孙佳亮. 论爬虫抓取数据行为的法律边界[J]. 电子知识产权, 2018(12): 58-67.
- [9] 孙晋, 冯涛. 数字时代数据抓取类不正当竞争纠纷的司法裁判检视[J]. 法律适用, 2022(6): 112-120.