

# 网络爬虫技术爬取数据的商业秘密保护

陈晓烨

华东政法大学 上海 200050

**摘要:** 大数据获取是大数据产业链中的重要环节,而网络爬虫技术是相对而言较为成熟的数据获取方式。在此背景下,企业数据的商业秘密属性受到了较大冲击。而对数据的商业秘密保护有助于维持利益平衡,促进知识经济发展。实际上,由于网络爬虫技术抓取的数据体量过于庞大,企业数据的秘密性并不天然丧失。且大体量的数据在互联网时代具有较强的分析价值,也需要足够的成本投入,具备足够的价值性。而保密性考察企业的保密意愿,而非保密措施的强度。因此,利用商业秘密保护被网络爬虫技术爬取的数据具有可行性。

**关键词:** 网络爬虫; 商业秘密; 价值性; 秘密性; 保密性

## 绪论

随着数字经济不断发展,企业愈发依赖数据进行生产活动,据预测,截止2025年,企业的数据利用率将达86%。<sup>[1]</sup> 相关数据能够帮助企业更了解用户兴趣以及行业前景研读,有助于企业及时调整经营政策,以便保持甚至扩大市场份额,具有较大的经济价值,在符合商业秘密秘密性、价值性以及保密性的前提下,企业可将其作为商业秘密进行保护。

而大数据产业下,不同网络平台之间的信息数据交互已然成为网络运作的核心,网络爬虫技术在查询、收集信息过程中的重要性日益凸显。该技术本质上属于一种计算机程序或脚本,能够按照程序编写者预设的触发条件,自动且高效地访问、下载、解析目标计算机信息系统中的数据,以便用户搜索的时候能够及时匹配,缩短向用户提供数据的时间。<sup>[2]</sup> 尽管网络爬虫技术能够帮助企业较为高效地获得相关数据,但同样增加了企业自身数据被挖掘的可能性。因此,有必要结合商业秘密保护的三要件,分析在此情况下,企业是否可以继续利用商业秘密制度对自身数据库进行较高强度的保护。

### 1. 网络爬虫技术对商业秘密的影响

商业秘密规定于《反不正当竞争法》第九条,该条款主要明确了侵犯商业秘密的行为以及商业秘密的概念。而《最高人民法院关于审理不正当竞争民事案件应用法律若干问题的解释》对于商业秘密的构成要件作了进一步的阐释,明确运用商业秘密进行保护的前提是保护对象满足价值性、秘密性以及保密性三个基本条件。产生价值性的前提是保护

对象具有秘密性,如果保护对象属于公共领域的内容,不具有秘密性,也当然不产生创造竞争优势的价值。当企业数据具备秘密性后,则需要分析该数据是否具备价值性,即对企业是否可以产生商业价值。在个案中,价值性的证明可以体现在被告是否使用或保护对象是否对被告产生价值,亦或是原告对该保护客体的投资成本,由于数据的价值主要以数据的质量为考量标准,不受爬虫技术爬取的影响,故关于数据的价值性将在后文统一分析。最后,则应当分析商业秘密的保密性,即企业所采取的保护措施的合理性。

#### 1.1 网络爬虫技术对保密性的冲击

秘密性的判定主要看其是否普遍知悉以及是否容易获得,简言之,是否存在较多人知晓的客观状态或者可能性。我国首个用户数据库作为商业秘密保护的案件<sup>[3]</sup>中,法院经审理认为涉案网站数据库整体不容易为相关领域的人员普遍知悉和容易获得,符合秘密性构成要件,且具有价值性与保密性,因而认定数据库属于商业秘密。

但成熟的网络爬虫技术帮助数据抓取方企业绕过常规的技术拦截方式来获取企业数据库中的信息,极大地冲击了相关数据的秘密性。该技术一定程度上提供了破解企业数据“非公开性”的可能,其所爬取的信息通常系互联网用户可以访问的公开信息。因此,在技术支持下,相关数据是否仍旧“不容易为相关领域人员普遍知悉和容易获得”就需要重新考量。

#### 1.1 网络爬虫技术对保密措施合理性的影响

为了使得网络爬虫技术能够继续发展,营造鼓励创新、

公平公正的良性竞争环境，2012年百度等互联网搜索引擎企业签署了《互联网那个搜索引擎服务自律公约》，规定不得在违反机器人协议的情况下爬取数据，以及不对机器人协议进行不当限制。该协议将爬虫排除标准<sup>[4]</sup>界定为国际通行的行业管理和商业规则。但爬虫排除标准是网站服务者自行设置的单方宣示，不具备法律意义上的强制执行力，其是否可以构成保密措施有待考量。而基于基于“数据包”、“访问行为”以及“流量统计和日志分析”的技术措施又可能不当地影响正常网段的使用<sup>[5]</sup>，反而对企业造成不利。因此，网络爬虫技术下，企业应当采取何种程度的保密措施也值得探讨。

## 2. 数据商业秘密保护利于维持各方平衡

尽管网络爬虫技术可能对企业的商业秘密保护带来一定影响，但却在一定程度上促进了相关数据信息的流通，为信息流通与社会进步提供了助力。因此，如何保护网络爬虫技术所抓取的数据这一问题实质是如何平衡企业、抓取方甚至公共利益之间的关系。需要明确的是，由于网络爬虫技术通常用于抓取大量数据，而这些数据通常是企业在用户授权后收集并汇总的，各方利用的是这些数据所蕴含的整体信息，因此此处的利益平衡考量将不过多地纳入用户个人的隐私权。

商业秘密的保护模式对于维持各方的利益平衡起到了关键作用。在这种模式下，通过合法手段获取的数据留在公共领域，供各方利用，促进信息的流通；而通过非法手段获取商业秘密则受到法律规制，保护企业的投入成本。对于“搭便车”行为和正常的信息传播交流进行有效区分，同样有助于市场区分搅局者，规范市场经营秩序，促进公平竞争。

同时，在互联网时代下，商业秘密的保护模式一定程度上更利于信息的披露。商业秘密通过法律保护形式取代了公司采用私力救济的方式，使得公司避免对保密措施进行过度投资，同时避免了企业在防范非法抓取行为的同时将合法抓取行为排除在外，妨碍用户正常上网，有利于创造性活动的进行。

值得注意的是，商业秘密的保护模式不仅能维持经营秩序，还有助于消费者以最小的成本获取足够的信息。该保护模式同样避免了版权法所产生的争议，对于知识经济的健康发展至关重要。通过避免信息的封闭，商业秘密保护模式使得企业能够更好地进行创新活动，促进技术迅猛发展，为

社会进步贡献力量。

总体而言，网络爬虫技术与商业秘密保护之间的平衡是一个综合性的问题，需要在法律、伦理和商业实践等多个层面进行权衡。商业秘密的保护模式不仅在信息自由与社会进步中发挥积极作用，还有助于维护良好的经营秩序，最终促进了知识经济的繁荣发展。

## 3. 网络爬虫技术下数据的商业秘密性分析

但适用商业秘密对网络爬虫技术抓取的数据进行保护存在着一定难度。商业秘密制度较为成熟的美国判例中对于通过网络爬虫爬取的数据是否侵犯商业秘密、侵犯作品汇编权或违反《计算机欺诈和滥用法》等方面有着多重的讨论。

<sup>[6]</sup>我国实务判例中，也出现法院认可数据构成商业秘密，但仍以《反不正当竞争法》一般条款进行判决的情形。<sup>[7]</sup>但相较于《反不正当竞争法》一般条款，商业秘密保护力度更大，更具优势，如果商业秘密保护规则存在适用的合理性与可行性，则应当优先适用于更为原则性的《反不正当竞争法》一般条款。<sup>[8]</sup>而在美国，也存在将被爬取的海量数据作为商业秘密进行保护的案例。<sup>[9]</sup>换言之，被爬取的数据存在满足商业秘密构成要件的可能。

### 3.1 被爬取的数据仍具有秘密性

网络爬虫抓取的数据可被用户所知悉是相关数据可否获得商业秘密保护的主要争论点，有观点认为，“公众可以自由获取”的信息容易获得，不可能构成商业秘密。而美国 *Compulife* 案的最终结果则反驳了该观点，法院指出，尽管抓取的单个报价不具有秘密性，但单个公众难以获得大量报价单，企业基于报价单形成的数据库仍然具有秘密性。美国《反不正当竞争法重述（第三次）》指出，所宣称商业秘密作为一个整体具有秘密性是决定性要素。商业秘密的部分或全部组成是众所周知的，并不排除对单个元素的组合、汇编或整合进行商业秘密保护。<sup>[10]</sup>而网络爬虫技术通常一次性抓取海量数据，被抓取的数据系企业经过长时间积累形成的海量数据，单靠相关领域普通技术人员获知难度巨大，因此，被抓取的海量数据的秘密性事实上根植于其规模性，故爬虫所爬取的海量数据能够满足秘密性的要求。

### 3.2 被爬取的数据具有价值性

商业秘密之所以收到保护是因为其具有价值性，能够为企业创造财富或竞争优势。大数据时代下，大数据对企业具有巨大的商业价值。比如北京淘友天下技术有限公司等与

北京微梦创科网络技术有限公司一案中，法院之所以认定用户数据构成微博商业秘密的原因，系用户数据信息对微梦公司具有重要的商业价值，可以在保持现有用户活跃度的同时持续吸引新用户。<sup>[7]</sup>而除了增强用户粘合力以及吸引新用户外，相关数据还可以被用以进行市场研究、消费者行为分析、创作优化、舆情监控以及行业预测。不难理解，这些海量的数据所形成的数据库在数字时代下是企业重要的生产资料，尽管价值性的判断通常在个案中进行，但由于被爬取的数据极容易被竞争对手用以研判并生产具有竞争性的产品或服务，其价值性较容易满足。

价值性同时还可以依据企业的投入成本进行认定，该观点在美国判例法中得到了普遍的承认<sup>[11]</sup>。我国司法实务中也认同原告的投入成本值得收到保护，北京阳光数据公司和上海霸才数据信息有限公司之间的合同纠纷案就是一个有效证明。

故网络爬虫技术所抓取的海量数据作为数据集合，较容易具备价值性。

### 3.3 企业保密措施仅需具备合理性

在分析企业保密措施的合理性之前，需要明确保密性的具体要求。在网络爬虫技术的辅助下，即使企业已经采取了充分的保密措施对商业秘密进行保护，但他人仍可能通过获得数据信息，在边际效益的考量下，过分严苛的技术措施性价比过低，反而不利于企业自身发展。故企业的保密意愿一定程度上有所降低。为避免企业保密意愿消失带来的不利后果，企业保密措施不应过于严苛。

事实上，商业秘密所要求的保密性仅仅是相对保密，而非绝对保密。保密性仅要求必须使得其他人在并未诉诸不法行为的情形下获取信息变得困难或者代价高昂。换言之保密信息仅需足以“防君子”即可，而不必防掉所有“小人”。如果保密措施需要严密到足以防止“防小人”才达到合理性之要求，则商业秘密条款的规定事实上不具备任何实施可能与意义。而互联网产业建立在互联互通以及信息的高利用率上，在这种互联互通的大背景下，要求企业设置高成本的技术保密措施，事实上违背了“理性经济人”的基本诉求。

在百度公司诉奇虎公司关于搜索引擎Robots协议案中，法院对于爬虫排除标准有了明确定位，认定该协议为互联网行业中重要的行业规则，属于搜索引擎行业内公认的、应该被遵守的商业道德。<sup>[12]</sup>则此类排除协议应当为市场所遵守。

尽管有观点认为，爬虫排除标准作为一种技术违反，并不能起到禁止访问的作用，无法构成《最高人民法院关于审理不正当竞争民事案件应用法律若干问题的解释》中所规定的保密措施。但美国司法实践中有判例指出，在特定情形下，爬虫协议可能构成网站访问进行有效控制的技术措施。换言之，爬虫排除标准并非当然排除于具有合理性的保密措施之外。

而美国 *Complife* 案<sup>[13]</sup>更进一步指出保密措施的合理性不应过于苛责，甚至说该要件是最容易达到的都不为过。该案原告创建了公开的网站允许用户免费申请保险报价，网站通过用户信息声称多个报价结果供用户选择。且该网站没有任何技术访问或合同的限制。而被告利用网络爬虫技术从原告网站上爬取了约 4300 万份报价，生成了约 80 万份独特的数据输入组合于自己网站上使用，并于原告开展竞争活动。法院认为，被告利用网络爬虫不当地爬取了过多的数据，已经超出了正常用户可获取数据的合理预期，即违背了原告关于正常用户不会大规模和自动化盗窃的合理预期。可见，网站不采取明显的技术措施限制用户并不等同于未能采取合理措施保护整个数据库。况且，企业可以从源头上对数据进行加密处理，使得程式化的爬虫技术收到更多干扰。

### 结论

随着数字经济不断发展，数据成为了重要的生产要素，其所承载的商业价值以及企业对其的投入均呈现上升趋势，如果不能对其予以合理的保护，则会极大打击互联网产业的发展积极性。而网络爬虫技术的应用，使得数据的商业秘密属性因爬虫抓取可公开访问而不具备秘密性的构成要件的质疑，且由于网络爬虫技术不断迭代，用户保护难度逐渐攀升，采取保密措施的意愿逐步下降。网络爬虫所爬取到的海量数据事实上影响了企业与公共利益的平衡、市场知识创造的积极性以及良好经营秩序，而采用商业秘密制度进行保护实现上述目标。

虽然互联网上各个用户均可轻易获得部分数据，但相较于网络爬虫技术所抓取的体量而言，单个用户所获取的体量微不足道，不足以破坏整体的秘密性。而这部分数据整体由企业搭建平台收集整理，可以帮助经营者增强用户粘合力以及进行市场判断研究，在互联网经济时代下具有较强的价值性。而商业秘密保护性仅要求保护措施达到相对保密的程度，因此爬虫排除标准网络爬虫技术领域作为已经形成的商

业道德,各经营者具备遵守的法律义务,可以满足合理性要求。故不应过分苛责企业的保密措施,应主要考量企业的保密意愿程度。

#### 参考文献:

- [1] 罗卓然.爬虫技术背景下用户数据库的商业秘密性研究[J].武汉冶金管理干部学院学报,2023,33(01):29-33.
- [2] 李峥妍.大数据时代下网络爬虫行为的刑法规制[J].市场周刊,2024,37(01):153-158.
- [3] 胡宓,玄玉宝.网站用户注册信息数据库的财产权归属及其性质[J].人民司法,2012(12):50-52. DOI:10.19684/j.cnki.1002-4603.2012.12.015.
- [4]Koster M. A method for Web robots control[EB/OL].(1996-12-04)[2024-02-15].<http://www.robotstxt.org/norobots-rfc.txt>
- [5] 黄克敏.网站信息安全之反爬虫策略[J].保密科学技术,2018(10):62-63.
- [6] Infogroup, Inc. v. Datasellc, 95 F. Supp. 3d 1170 (D. Neb. 2015).
- [7] 北京知识产权法院.北京淘友天下技术有限公司等与北京微梦创科网络技术有限公司不正当竞争纠纷二审民事判决书(2016)京73民588号.
- [8] 杨雄文,黄苑辉.论大数据的商业秘密保护——以新浪微博诉脉脉不正当竞争案为视角[J].重庆工商大学学报(社会科学版),2019,36(04):138-145.
- [9] Compulife Software Inc. v. Newman, 959 F.3d 1288 (11th Cir. 2020).
- [10] Restatement (Third) of Unfair Competition § 39 comment f (1995).
- [11] MAI Systems Corp. v. Peak Computer, Inc., 991 F.2d 511 (9th Cir. 1993).
- [12] 北京市第一中级人民法院.北京百度网讯科技有限公司、百度在线网络技术(北京)有限公司与北京奇虎科技有限公司不正当竞争纠纷案(2013)一中民初字第2668号。
- [13] Compulife Software Inc. v. Newman, 959 F.3d 1288 (11th Cir. 2020).

#### 作者简介:

陈晓焯(2000.9-),女,汉族,福建福州人,硕士研究生,研究方向:知识产权。