doi:10.13582/j.cnki.1672-7835.2021.04.018

滥用网络爬虫技术收集个人信息的刑法规制

宋行健
(中国政法大学 刑事司法学院,北京 100088)

摘　要:随着网络爬虫技术的信息收集能力不断提升,其在被滥用时所造成的损失日益严重。认定“非法收集”个人信息,不仅应考察方法本身是否具有非法性质,还应考察收集信息的依据或资格。对突破反爬虫技术措施收集个人信息的行为,应根据法条之间的竞合关系,适用侵犯公民个人信息罪。对正常登录系统收集个人信息的行为,应考察收集行为是否出于履职的必要。对收集网站外部访问者的个人信息的行为,应根据网站功能、浏览内容确定信息的收集权限。

关键词:网络爬虫;个人信息;电子数据;非法收集;最小必要
中图分类号:D924.3　　**文献标志码:**A　　**文章编号:**1672-7835(2021)04-0139-10

一　问题的提出

互联网 3.0 时代的到来,不同网络平台之间的信息交互已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。网络爬虫在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①。这一技术虽然在精准搜索、舆情监控、系统漏洞检测、大数据挖掘等领域得到了广泛运用,但由于相关的技术资源易于获取,技术原理易于掌握,当使用者为了非法获取个人信息而滥用这一技术时,将造成公民个人信息的大范围泄露。例如在 2020 年 11 月由媒体报道的“圆通快递泄露 40 万条公民个人信息”事件中,该公司的 3 名员工擅自将快递员账号有偿出租给外部人员,这些外部人员通过网络爬虫技术查询、导出快递信息,其中包括收件人与发件人的姓名、电话、地址,随后又打包出售给实施电信诈骗的团伙②。

纵观国内有关滥用网络爬虫技术收集个人信息的研究,主要集中在对入罪范围的探讨。第一种观点认为,应当构建形式入罪、实质出罪的机制,形式上应遵守相关法律关于公民个人信息的保护规则、用于规范爬虫行为的行业规则,实质上应在权限许可范围内获取公民个人信息,或是仅收集无法与特定个体的身份相对应的信息③。第二种观点认为,应从行为不法、对象不法这两个层面探讨网络爬虫的刑事违法性,在行为人突破了反爬虫技术措施或使用网络爬虫取得了限制访问、获取的数据时,应承担刑事责任④。第三种观点认为,应考察收集行为是否具有合法性、正当性与必要性,以及行为人的主观认识、网络爬虫的功能、滥用网络爬虫技术所造成的结果,对入罪范围予以限制⑤。第四种观点认为,应当从访问权限、

收稿日期:2021-01-20
基金项目:最高人民法院执行研究重点课题(2018ZGFYZXKT201829)
作者简介:宋行健(1995—),男,湖南衡阳人,博士生,主要从事刑法学研究。

①游涛,计莉卉:《使用网络爬虫获取数据行为的刑事责任认定——以“晟品公司”非法获取计算机信息系统数据罪为视角》,《法律适用》2019 年第 10 期。
②《圆通多位“内鬼”有偿租借员工账号,40 万条公民个人信息被泄露》,《新京报》2020 年 11 月 16 日。
③刘艳红:《网络爬虫行为的刑事规制研究——以侵犯公民个人信息犯罪为视角》,《政治与法律》2019 年第 11 期。
④杨志琼:《数据时代网络爬虫的刑法规制》,《比较法研究》2020 年第 4 期。
⑤张一献:《从技术到犯罪:恶意网络爬虫行为入罪的类型认定与裁判思路探索》,《时代法学》2020 年第 4 期。

数据性质、合理使用义务这三个角度,厘清网络爬虫的犯罪边界、数据性质对罪名适用的影响、非法获取计算机信息系统数据罪与侵犯公民个人信息罪可能存在的竞合①等问题。

然而,学界现有的研究成果主要是基于突破反爬虫技术措施收集个人信息而展开,既未总结司法实践中的行为样态,类型化、场景化地对非法收集个人信息的司法认定问题逐一分析,又未结合2021年4月发布的《中华人民共和国个人信息保护法(草案二次审议稿)》(以下简称《个人信息保护法》草案二审稿)中关于个人信息保护的基本原则,在此基础上结合刑法的规范目的,探讨收集信息行为的刑事违法性依据。具体而言,有以下几个值得进一步考察的问题:第一,立足于个人信息在不同法律领域中的保护方式,在滥用网络爬虫技术收集个人信息的情况下,为何具有运用刑法予以规制的必要性。第二,在司法实践中,滥用网络爬虫技术非法收集个人信息有哪几种表现形式,因为这些表现形式不仅体现了网络爬虫技术的不同运作方式,而且体现了不同的刑法规制侧重点。第三,在这些表现形式之下,应如何结合对侵犯公民个人信息罪构成要件的解释,对信息收集行为的刑事违法性予以认定。本文将围绕以上三个方面的问题逐一展开研究。

二　刑法规制的必要性分析

(一)个人信息在不同法律领域中的保护方式

其一,私法从个人信息所适用的场合出发,确定个人信息所对应的具体法益内容②。例如《中华人民共和国民法典》(以下简称《民法典》)立足于个人信息与隐私权之间的关联性,以及信息在流转过程中所产生的商业价值,以事前赋权的方式,对滥用网络爬虫技术的侵权属性予以确认。《民法典》在第六章中一并规定了涉及"隐私权"与"个人信息"的保护措施,二者之间的交集体现为个人信息中的私密信息。由此可见,个人信息权与隐私权存在客体上的重合③。由于隐私权归属于人格权之中,是作为主体的人能够生存、发展的必要条件,根据人身权益优先保护的原则,隐私权具有优先性的法价值。因此《民法典》第1 034条指出,个人信息中的私密信息,应当优先适用与隐私权相关的规定,而非与个人信息保护相关的规定。又如《中华人民共和国反不正当竞争法》第9条对一系列侵犯商业秘密的行为予以禁止,这旨在保护竞争秩序以及经营者、消费者的合法权益。而在使用网络爬虫技术获取商业秘密的过程中,使用者也可能对公民个人信息造成侵犯。例如"客户名单"这类商业秘密包含着较为丰富的信息类型,其记录了客户名称与联系方式、收货地址、对产品的需求、交易习惯等综合性信息,这些信息能够为经营者创造商业利益,受到保密措施的严格保护,很难从公开途径直接获取④。若这些信息被竞争对手利用网络爬虫技术获取,不仅会使拥有该商业秘密的经营者遭受经济损失,同时也违背了个人信息处理的合法、正当、必要原则。

其二,公法则主要针对个人信息的处理过程予以规制,规制内容包含个人信息的获取、存储、使用等一系列行为,根据信息的类别与数量确定处罚规则,而非将重点放在信息所反映的某类具体法益上。在公民个人信息领域,刑法早于其他部门法明确了入罪的界限,不仅通过侵犯公民个人信息罪对非法获取、提供个人信息的行为予以处罚,而且通过兜底性的非法获取计算机信息系统数据罪,为网络爬虫所可能造成的各类法益侵害留下了足够的惩治空间⑤。一方面,即使是在侵犯公民个人信息罪中,也包含着多样化的法益内容,不仅包括公民个人信息的安全和自由,也包括与这些信息相关的社会管理制度、社会对这些信息的处置方式⑥。另一方面,数据拥有比信息更为广泛的外延,是信息在网络空间中予以展示的必要载体。因此,数据基于其公共性原理,具有

①付强,李涛:《网络爬虫的刑法应对》,《中国检察官》2020年第18期。

②程啸:《论大数据时代的个人数据权利》,《中国社会科学》2018年第3期。

③杨卓黎:《民法典背景下新型人格权评析及其保护——基于人格权体系协调的立场》,《湖南科技大学学报(社会科学版)》2019年第3期。

④何炼红:《网络公司客户名单商业秘密与个人信息隐私权的冲突与协调》,《法学杂志》2010年第12期。

⑤杨志琼:《非法获取计算机信息系统数据罪"口袋化"的实证分析及其处理路径》,《法学评论》2018年第6期。

⑥凌萍萍,焦冶:《侵犯公民个人信息罪的刑法法益重析》,《苏州大学学报(哲学社会科学版)》2017年第6期。

承载传统法益、新型法益的多样化功能,未受到某类信息所体现的特定法益内容的限制。随着各类法益逐渐能够通过数据的形式进行传输和存储,刑法也更能适应对各类法益全面保护的需求。

(二)严重而广泛的侵害后果需要刑法介入

首先,网络爬虫技术与人工智能相结合,其所具备的信息识别、收集能力不断提升,在被滥用时所造成的损失日益严重。在网络爬虫技术的应用过程中引入算法,能够提升获取、加工信息的效率,但公众的信息安全、网络空间秩序也随之受到挑战。网络爬虫除了按照使用者预设的条件自动筛选、收集、下载数据,从而使个人信息面临泄露的风险外,还会为了在一定时间内访问更多的信息,占用大量的网络带宽与硬件资源并产生巨大的扫描流量,造成被收集信息的网站出现服务器过载甚至崩溃的严重后果,网站经营者也将因此遭受损失。

其次,网络爬虫的信息收集来源逐步拓展,在被滥用情况下造成的损失具有广泛性。网络爬虫收集信息的来源由电脑网页、计算机数据库向着智能手机等移动设备不断延伸①。由于各类电子设备、信息平台承载着不同的社会功能,因此网络爬虫所收集的个人信息类别也将更为多样,行为人能够借助不同的信息渠道,全方位地了解被害人的情况,在实施诈骗等犯罪行为时也能形成更强的针对性。此外,为了适应不同设备与信息源上的信息存储形式,网络爬虫的运作方式日趋复杂、隐蔽,即使信息的管理者设置了反爬虫技术措施,也只是降低了网络爬虫收集信息的效率,无法从根本上实现对滥用网络爬虫现象的有效遏制。

最后,通过滥用网络爬虫技术所收集的个人信息,能够使下游的其他犯罪更为便捷地实施。例如在非法放贷的过程中,放贷者利用网络爬虫技术侵入借款人的智能手机,非法收集其中存储的通讯录、聊天信息、行踪轨迹等内容,将这些信息提供给职业化的催收团伙,由他们通过电话骚扰借款人的亲友、上门滋扰等方式,不断向借款人施加心理压力。当放贷者形成网络套路贷的经营模式之后,能够长期地以此谋取非法利益。又如行为人非法利用“深度伪造”技术,先利用网络爬虫技术收集被害人公开发布的照片、视频,从中提取被害人的人脸信息,交由计算机程序进行深度学习,再使用前述程序篡改其他的真实视频,将其中的原始信息替换为被害人的人脸信息,从而生成真假难辨的新视频,用于实施诈骗等犯罪活动②。

三　滥用网络爬虫技术收集个人信息的行为样态

(一)违反爬虫协议或突破技术措施收集个人信息

根据网站管理者对网络爬虫所采取的应对措施,可以划分为协议、技术两个层面的内容。首先在协议层面,包括外部公约与内部协议。“外部公约”是指与网络爬虫应用领域相关的从业者,在达成共识的情况下制定的自律性公约或行业指引性规范,例如《互联网搜索引擎服务自律公约》《App 违法违规收集使用个人信息自评估指南》。这些规范能够在不同类型的应用场景下,为网络爬虫的使用者提供具体的指引,具有较强的可操作性。“内部协议”是指网站管理者根据网站中存储信息的公开程度,按照自身意愿对信息进行分类,要求网络爬虫只能在网站管理者设定的范围内收集信息。这类协议既能以文字的形式向网络爬虫的使用者展示,又能以代码的形式向网络爬虫展示。前者通常采取点击生效或浏览生效的方式,后者则是在网站根目录下,以文本文件的形式设置爬虫协议。“内部协议”由于具有简单高效的特性,因此成了互联网行业普遍遵循的技术规范,但其并未在技术层面通过强制性的措施制止网络爬虫对数据的获取。其次在技术层面,网站管理者往往通过各类反爬虫技术措施,以内外相互独立的两种机制加强对个人信息的保护。内部机制体现为网站管理者通过账号、密码进行身份认证,确认用户具备访问计算机信息系统的权限。外部机制则是通过分析访问者的 IP 地址、访问频率,在智能筛选的基础上有针对性地屏蔽一部分疑似网络爬虫的访问请求,这在技术上表现为 IP 限制、核对验证码、参数签名等③。值得探讨的问题是,在协议、技术两个不同

①曹阳:《我国对违反“爬虫协议”行为的法律规制研究》,《江苏社会科学》2019 年第 3 期。

②李怀胜:《滥用个人生物识别信息的刑事制裁思路——以人工智能“深度伪造”为例》,《政法论坛》2020 年第 4 期。

③贺思聪:《爬虫实战:从数据到产品》,电子工业出版社 2019 年版,第 5 页。

的层面,是否会对收集个人信息行为的刑事违法性认定产生不同的影响?

(二)正常登录系统非法收集个人信息

行为人在拥有单位提供的账号、密码的情况下,无需通过侵入、破坏的手段破解反爬虫技术措施,就能便捷地访问单位内部网站,在登录之后使用网络爬虫收集由单位所保存的非公开的个人信息。根据司法实践中相关单位保存个人信息的原因,可分为三种情形:第一,基于隶属关系。如由单位的信息管理系统存储的、在单位任职的员工信息。在利用网络爬虫技术擅自收集这些信息之后,行为人不仅能够了解到单位内部各层级的人员组成情况,而且侵犯了相应员工的个人信息权。第二,基于业务关系。例如公司在与客户开展交易的过程中,能够获取客户的联系方式、家庭住址等个人信息,当它们被网络爬虫非法收集之后,相关的产品营销公司能够持续地向客户开展推广、宣传等活动,从而对客户的生活造成负面影响。第三,基于履行法定职责。即由国家机关在履职过程中所存储的公民个人信息,例如工商行政管理部门所存储的个人工商登记信息,包含着法定代表人的姓名以及注册时使用的移动电话号码①。在这类情形下,值得进一步探讨的是:如果行为人在履职过程中能够通过正常渠道获取个人信息,相较于运用网络爬虫技术所达到的信息获取效果,二者是否存在本质的区别?

(三)非法收集外部访问者的个人信息

行为人在发现网站技术漏洞的情况下,能够通过在网站内写入代码的方式使用网络爬虫收集来自外部访问者的个人信息。此时,网络爬虫所收集的信息类型与网站所具备的功能密切相关,体现了外部访问者对于特定商品或服务的需求。例如行为人利用医院网站的管理漏洞,擅自写入网络爬虫代码,当访问者浏览医院网站上的内容,或是就病情进行网上咨询时,通过设定"病患"等关键词,自动收集访问者用于上网的设备信息、健康生理信息,并将它们存储到行为人所管理的数据库中,有偿提供给需要推广业务的医院②。由于网站管理者针对网络爬虫所设置的协议、技术,是为了对来自外界的访问进行筛选,从而保护网站所存储信息的安全,因此在网络爬虫仅收集外部访问者个人信息的情况下,其刑事违法性具有单独探讨的必要性。第一,如果网站也对访问者的信息予以收集,而且已经向访问者就收集范围作出明示,网络爬虫的使用者在得知后,也按照前述范围收集访问者的信息,能否认定其行为的刑事违法性?第二,对于访问者所产生的网页访问记录、网页浏览时长等"数位足迹"信息,是否应排除在个人信息的认定范围之外?

(四)三种行为样态在刑法规制中的不同侧重点

滥用网络爬虫技术收集个人信息,在司法实践中涉及三种行为样态,它们都需要结合《中华人民共和国刑法》第253条的"侵犯公民个人信息罪"予以认定,围绕构成要件的解释、违法程度的认定分别展开,但涉及的侧重点各有不同。其一,在"违反爬虫协议或突破技术措施收集个人信息"的情形下,获取个人信息方式的违法性、破坏性较为明显,与侵犯公民个人信息罪的罪状中的"窃取或者以其他方法非法获取公民个人信息"具有较高的契合度,但其民刑界限仍有待进一步区分。其二,在"正常登录系统非法收集个人信息"的情形下,获取方式本身的非法性较弱,因为行为人收集的个人信息,与行为人在正常履行职责过程中所能接触到的个人信息具有一定的关联。但在该情形下,有必要结合行为人收集个人信息的范围、数量,探讨其是否具有收集的依据或资格,是否出于履行职责的必要,进而认定收集行为是否需要予以刑法规制。其三,在"非法收集外部访问者的个人信息"的情形下,网络爬虫所收集个人信息的来源与前两种情形存在区别,即并非来源于特定的网站或平台,而是来源于不特定的多名网站访问者,因此能够收集类型更为多样的信息,主要涉及刑法意义上的"个人信息"的外延认定问题以及收集个人信息的依据问题。

四　各类行为样态的刑法规制

(一)"非法收集个人信息"的总体认定依据

侵犯公民个人信息罪将"非法获取公民个人

①北京市通州区人民法院(2019)京0112刑初62号刑事判决书;沈阳市经济技术开发区人民法院(2018)辽0191刑初418号刑事判决书。

②苏州市中级人民法院(2018)苏05刑终1142号刑事裁定书。

信息”的行为纳入惩治的范围,“非法获取”的途径包括“窃取或者以其他方法”,“窃取”是指秘密的或不为人知的方法①。在前文所探讨的三种情形中,“正常登录系统非法收集个人信息”“非法收集外部访问者的个人信息”可以被评价为以“窃取”的方法,而在“违反爬虫协议或突破技术措施收集个人信息”的情形下,则不能排除管理个人信息的主体知情的可能性,因此还需要结合本罪罪状中“其他方法”的含义进行判断。

通过对相关指导案例裁判观点的归纳,本罪中的“其他方法”不需要与明文规定的“窃取”具有相当的违法性,可分为以下三类。第一,以本身带有违法性或破坏性的方法获取,例如骗取、侵入计算机信息系统等。通过这些方法一般获取的是“第一手”信息,可能导致更多本身不在“流通”市场的公民个人信息被转卖或传播,其产生的危害性更大②。第二,未获得授权而获取,此时行为人无权了解、接触相关公民的个人信息。第三,以不正当方式获取,此时行为人违背了信息所有者的意愿或真实的意思表示,或者违反了社会公序良俗③。

由此可见,“其他方法”在司法实践中的表现形式具有广泛性。由于“非法获取”与“非法提供”并列,二者同属于侵犯公民个人信息罪的表现形式,因此,“非法获取”的认定不以信息的具体用途为依据,关键在于获取行为本身。有学者指出,除了应考察这些方法本身是否具备非法的性质,还应当考察行为人是否具有获取个人信息的法律依据或资格④。《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》(以下简称《个人信息刑案解释》)将“收集”作为“获取”的下位概念,但仅在第 4 条将在“履行职责、提供服务”过程中违反国家有关规定的收集行为,明确纳入“其他方法”的范畴。在网络爬虫技术不断发展的时代背景下,收集个人信息不再需要依托特定的职责或服务内容,因此仍有必要研究其他收集行为的刑事处罚边界。

基于体系解释的原理,对侵犯公民个人信息罪中的“非法获取公民个人信息”,应当以是否违反国家有关规定作为判断标准,即法律、行政法规、部门规章有关公民个人信息保护的规定⑤。这一判断标准也符合法秩序统一原理。一方面,这一原理能够在一定高度上确保不同部门法的保护目的不发生冲突,增强法律体系的协调性。刑法在设定处罚规则时,应当体现其保障法的地位与“二次规范”的性质,需要以其他部门法中已有的规制措施作为前置条件。另一方面,在理解刑法中“违法性”的含义时,应坚持违法判断的相对独立性,以刑法特有的价值追求与规范目的,确立独立解释的品格,但也有必要立足于其他部门法领域的视角,为刑事违法性的判断提供支撑⑥。诸如《中华人民共和国网络安全法》《信息安全技术个人信息安全规范》《关于加强网络信息保护的决定》都明确了个人信息收集应遵循合法性原则,收集个人信息的范围以服务提供者的服务内容、双方的具体约定为限。《中华人民共和国网络安全法》更是在第 27 条直接对滥用网络爬虫的行为予以禁止,具体包括非法侵入网络、窃取网络数据、干扰网络功能这三类行为,开发者既不能向他人提供具备以上功能的程序、工具,也不能利用这些程序、工具开展危害网络安全的非法活动。《个人信息保护法》草案二审稿则采取抽象与具体相结合的方式,首先在第 5 条确立了合法性原则,接着又在第 6 条至第 9 条分别确立了最小必要、公开透明、完整准确处理等各项具体的原则,为认定信息处理行为的合法性提供了参考。

虽然使用网络爬虫技术多是为了获取非公开的信息,但行为人也可能仅为提高获取信息的效率,利用网络爬虫技术收集已在网络上公开的个人信息,而且这些信息的公开已经过相应个体的

①沈德咏:《刑法修正案(九)条文及配套司法解释理解与适用》,人民法院出版社 2015 年版,第 188 页。

②苏琼,余丹:《周娟等非法获取公民个人信息案——非法获取大量公民个人信息的行为,如何定罪量刑》,《刑事审判参考》2011 年第 4 期。

③钟莉,范冬明:《胡某等非法获取公民个人信息案——通过非法跟踪他人行踪所获取的公民日常活动信息是否属于公民个人信息?》,《刑事审判参考》2014 年第 4 期。

④赵秉志:《公民个人信息刑法保护问题研究》,《华东政法大学学报》2014 年第 1 期。

⑤周加海,邹涛,喻海松:《〈关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释〉的理解与适用》,《人民司法(应用)》2017 年第 19 期。

⑥曾根威彦:《刑法学基础》,黎宏译,法律出版社 2005 年版,第 214 页。

同意。本文认为,此种情形应属于刑法保护的例外,理由如下:其一,应考虑到信息流转、使用的正常需求。例如企业在经营的过程中应公开一系列信息,其中就包括经营者的姓名与联系方式,以便于消费者查询、识别,这属于个人为了企业经营的需要,对个人信息权作出让渡、牺牲,使个人信息从属于企业信息①。第二,公民将个人信息予以公开,属于对信息的利用方式之一。例如公民将个人信息公开用于商贸、广告等用途时,能够使相关的商品或服务被更多的潜在客户所知悉。在行使个人信息权的过程中,公民出于趋利避害的考量,在确定公开个人信息的范围、方式时,都会作出一定的限制,例如只选择性地公开对个人隐私、生活影响较小的信息,从而确保它们在被非法使用的情况下,所引发的潜在风险处于可以预见、控制的范围之内②。因此,公民公开敏感隐私信息之外的一般信息,具有被害人承诺的性质,使得运用网络爬虫收集这些信息的行为能够阻却违法性。第三,网站管理者对于已经在网站上公开的个人信息,不能再从协议、技术层面对网络爬虫的收集行为要求进一步的授权。网络爬虫能够随时访问、收集这些信息,而且未改变信息的存在状态,例如由限制访问的状态转变为开放的状态,因此其在手段、结果两方面都不具有非法性。第四,有必要将获取、使用已公开的个人信息予以区别对待,当行为人将获取的信息用于非法活动时,仍然可以适用诈骗、寻衅滋事等相关罪名予以惩治。《个人信息保护法》草案二审稿第28条指出,对于已公开的个人信息的处理,应当符合信息被公开时的用途,不得超出与之相关的合理范围,否则应取得相应个体的同意。换言之,公民在将个人信息予以公开之后,只是希望潜在的目标对象能够自主获取、合法使用,未同意对这些信息不加限制地予以利用。此时,应当将个人信息的保护利益置于比利用利益更为优先的地位③。

(二)违反爬虫协议、突破技术措施时的认定

在网站管理者从协议、技术两个层面对网络爬虫予以限制时,刑法有必要将后者作为"非法收集"的判断标准。从爬虫协议的性质来看,网络爬虫的使用者如果能够在未触发反爬虫技术措施的情况下,对网站上公开的信息进行查询、收集,那就意味着网站管理者认同了网络爬虫的信息收集行为,并与网络爬虫的使用者形成了信息服务合同关系,这主要体现在搜索引擎的运作过程中④。因此,爬虫协议不仅是一种能够被网络爬虫识别的技术工具,还是前述信息服务合同中的重要内容,它以格式条款的形式,确立了网络爬虫收集信息的范围。如果网络爬虫的使用者无视其限制,则应当对擅自收集个人信息的行为承担相应的违约责任。相较而言,如果网络爬虫使用者在违反爬虫协议的基础上,进一步突破反爬虫技术措施收集个人信息,产生了更为严重的危害,应认定为刑法意义上的"非法收集",其理由有以下几个方面。

首先,反爬虫技术措施以代码的形式构建了个人信息的保护屏障,既为网站上的信息划分了不同开放程度的空间,又对访问者设置了不可逾越的权限范围,能够被认定为计算机信息系统的组成部分。根据《计算机信息系统安全保护条例》的规定,"计算机信息系统"包含计算机以及相关的设备、设施以及网络,其表现形式不以物理形态为限,而且信息安全是前述系统安全保护工作中的一项重要内容。非法侵入计算机信息系统罪中的"侵入",其实质是违背他人的意愿,在未经授权的情况下进入他人的计算机信息系统,例如采取技术手段突破安防设置并强行进入⑤。网络爬虫的使用者在故意规避、突破技术措施的情况下,能够使相关的安全策略陷入无效的状态,属于该罪中的"侵入"。在侵入前述系统并获取数据的情况下,与数据安全相关的保密性、完整性与可利用性等法益将随之受到侵犯⑥。

其次,当公民个人信息以数据的形式展现时,非法获取行为同时也符合侵犯公民个人信息罪、非法获取计算机信息系统数据罪的构成要件,两

①吴心斌,温锦资:《公民个人信息刑法保护的例外》,《人民法院报》2018年6月21日。
②张新宝:《从隐私到个人信息:利益再衡量的理论与制度安排》,《中国法学》2015年第3期。
③董悦:《公民个人信息分类保护的刑法模式构建》,《大连理工大学学报(社会科学版)》2020年第2期。
④宁立志,王德夫:《"爬虫协议"的定性及其竞争法分析》,《江西社会科学》2016年第1期。
⑤郎胜:《中华人民共和国刑法释义》,法律出版社2015年版,第491页。
⑥徐育安:《资讯风险与刑事立法》,《台北大学法学论丛》2014年第9期。

罪之间属于特殊法与普通法的关系，应按照法条竞合的原则以前罪论处。这是由于在网络空间中，“公民个人信息”与“计算机信息系统数据”存在概念上的交叉关系。相较于信息而言，数据的内容更为繁复、外延更为广泛，是包含了信息与其他数据冗余的集合。信息的数字化是其能够在网络空间中传输、存储的必要条件，因此信息作为数据经过加工处理之后的可视化内容，始终需要以数据为载体。退而言之，当网络爬虫收集的并非公民个人信息，但在获取手段上具有侵入性质的情况下，可将其认定为非法获取计算机信息系统数据的性质①。

最后，国外学界有关网络爬虫的入罪标准也呈现出缓和的趋势。一方面，承担刑事责任的关键判断因素由合同责任转变为技术责任，限制与缩小了使用网络爬虫技术的入罪范围。根据“代码理论”，应当以行为人在使用网络爬虫的过程中，是否故意规避、突破了网站管理者设置的技术措施作为民事责任与刑事责任的界限②。另一方面，应当对网络爬虫所收集的信息进行甄别、筛选，遵循“明确分割原则”。对于那些向访问者公开的、未通过技术措施予以保护的内容，以及未侵犯个人信息权、知识产权的内容，不应作为刑事责任的认定依据③。

（三）正常登录系统收集个人信息时的认定

如果行为人利用工作便利，在未突破反爬虫技术措施的情况下，通过合法使用单位提供的账号、密码登录内部网站，获取由单位保存的员工、客户、行政相对人等个体的信息，则需要考察行为人的收集行为是否出于履职的必要。首先，这有利于对形式上属于合法登录，但实质上属于“侵入”的行为予以准确定性。如果行为人已从单位离职，在单位未及时注销其登录权限的情况下，继续使用任职期间掌握的账户、密码登录单位内部网站，或是在履职期间擅自将登录信息告知未在单位工作的其他人员，使得他们能够利用网络爬虫收集单位存储的个人信息，此种行为虽然没有触发网站管理者设置的反爬虫技术措施，但由于未经单位授权，而且这些信息收集者不具备履行职责的条件，因此仍然属于侵入计算机信息系统非法收集个人信息。其次，如果行为人在履行职责的过程中，需要根据具体的工作内容获取个人信息，而且获取信息的行为符合单位的内部管理规定，则不具有非法性。反之，如果其超出了单位内部管理规定的范围，出于履职之外的需要而获取了个人信息，则无论其是通过反复记忆、私下抄录、手动复制等方式获取，还是使用网络爬虫技术收集，二者只是在非法获取信息的效率上存在区别，均应根据获取信息的数量、信息的类别，判断是否具有运用刑法予以规制的必要。

值得进一步讨论的问题是，如何认定单位内部规章制度与刑法意义上“非法获取公民个人信息”之间的关联性？如前所述，侵犯公民个人信息罪中的“非法获取公民个人信息”，需要结合国家有关规定进行认定。在该情形下，需要结合《民法典》与《个人信息保护法》草案二审稿中确立的基本原则，从行为人、单位、信息对应的特定个体之间的关系展开。

其一，单位内部规章制度包含了与个人信息相关的保护措施，以及个人信息的获取、存储等规则。根据《中华人民共和国网络安全法》第 17 条的规定，网络运营者应当履行安全保护义务，以防止未经授权的访问造成数据的泄露，这些措施包括内部安全管理制度的建设、数据的分类与加密等。因此，单位内部规章制度中的前述内容，应当得到网络爬虫使用者的遵守。当单位招录员工之后，会通过多种途径确保员工知晓与个人信息保护相关的内部规章制度，例如统一组织培训与考核，或在劳动合同中明确约定个人信息的保密要求。此外，单位为了凸显这些个人信息的重要性，往往会将其列为敏感信息，规定比其他信息更高的保密等级，对这些信息的提取、下载需要经过专门的授权，当员工擅自实施时将会受到单位内部的处罚。从员工、客户、行政相对人的角度而言，他们对于单位存储其个人信息的情况是知情的，未授权单位向其他主体提供这些个人信息，并且

①林维：《数据爬取行为的刑事司法认定》，《人民检察》2020 年第 4 期。

②Myra F. Din.“Breaching and Entering: When Data Scraping Should Be a Federal Computer Hacking Crime”, *Brooklyn Law Review*, 2015 (1): 410.

③Amber Zamora.“Making Room for Big Data: Web Scraping and an Afirmative Right to Access Publicly Available Information Online”, *Journal of Business, Entrepreneurship & the Law*, 2019(2): 203.

能通过查阅单位有关个人信息保护的规定,了解到单位对个人信息的使用方式。这符合《个人信息保护法》草案二审稿第14条确立的"知情同意原则",即特定个体应当在充分知情的情况下,自愿且明确地同意他人处理个人信息。《民法典》第1 035条进一步指出,处理个人信息应当以特定个体同意的范围为合理限度。因此,单位享有合法存储、使用个人信息的资格,而网络爬虫则能够在未经授权的情况下,批量获取符合预设条件的信息,例如单位员工的姓名、联系方式与住址。这一收集信息的行为,既未经过与信息相对应的个体的同意,也并非出于履行法定职责、维护公共利益等合理目的,因此具有非法性。

其二,单位内部规章制度不仅规定了与个人信息保护相关的措施,而且规定了不同职位获取个人信息的权限。《个人信息保护法》草案二审稿第6条指出,除了需要具备明确、合理的个人信息处理目的,在实现这一目的的过程中,还应在必要的最小范围内处理个人信息,采取对个人权益影响最小的方式,将与处理目的缺乏关联的个人信息排除在外。因此,行为人只有在遵守单位内部规章制度的情况下,才能确保其处理个人信息的行为处于履行职责的必要限度内。一方面,行为人在履行职责过程中查询、获取的个人信息,其范围具有局限性,不能直接按照指定条件高效、全面地收集信息。例如当公司员工仅负责某一地区的业务时,其在工作过程中对于客户信息的浏览权限将受到区域范围的限制,不能任意地浏览公司所拥有的全部客户信息。又如快递公司所设置的物流风险控制系统,能够监测快递员是否有违规异地查询其他网点运单号信息的行为。另一方面,在行为人超越职权范围的情况下,还应结合信息对于特定个体身份的识别功能,判断收集行为是否具有刑事违法性。例如在工商行政管理部门存储的信息中,既包括个体工商户和企业的备案登记信息、行政许可与行政处罚信息、经营者的姓名与联系方式这些属于应当公开的信息,又包括前述主体在备案登记的过程中所提交的经营者的身份信息等,这些信息不在对外公开的范围内。因此,如果行为人在登录单位系统之后,擅自对后一类信息使用网络爬虫技术予以收集,则可认定为非法获取公民个人信息的性质。

例如在"余某侵犯公民个人信息案"中,被告人余某于2014年4月至6月在某电商公司工作期间,先登录其在公司内部论坛网站的账号,再使用自行编写的网络爬虫程序,擅自收集该公司员工的个人信息两万余条,并将这些个人信息存储于电脑硬盘之中,在离职之后带走。余某辩称只是为了了解公司的组织架构,但法院认为前述收集行为具有窃取公民个人信息的性质,理由在于:第一,该公司出于保护个人信息的需要,没有统一编制通讯录,员工只能出于工作的必要查询相关的员工信息,而且使用这些敏感信息必须经过授权。根据该公司的数据安全规范,使用信息的方式包括"直接使用",即通过网络爬虫技术进行信息的提取与挖掘,前述行为必须经过授权,员工不得擅自大批量提取、挖掘个人信息,即使这些信息对内是可以查询、浏览的。第二,余某在任职期间,明知公司未提供通讯录,不可擅自获取全部员工信息,却故意使用网络爬虫技术,通过自己在单位的账户以及在系统内部有针对性地查询、浏览的权限,在相关信息的所有者、保管者、经手者即公司和员工个人均不知情亦未发觉的情况下,将大量的员工个人信息秘密地爬取、保存并在离职时带走,因此其行为具备窃取公民个人信息的性质①。

(四)收集外部访问者个人信息时的认定

如果网站也对外部访问者的个人信息予以收集,而且已就收集范围作出明示,网络爬虫的使用者在得知后,也按照前述范围收集访问者的个人信息,能否认定为刑法意义上的"非法收集",本文对此持肯定观点。个人信息保护领域的"场景理论"认为,信息的保护与流转需要以特定的场景为依托,应当根据不同的应用场景,对信息安全采取差异化的保障措施,如此才能符合各方的预期②。在信息的处理过程中,如果处理者将信息从特定的场景中抽离,转换到信息主体不了解、不熟悉的场景中,或是将他人的零散信息擅自予以组合,则违背了"场景正义",将使他人的正常生活遭受困扰。因此,即使访问者同意网站收集其

①杭州市余杭区人民法院(2017)浙0110刑初737号刑事判决书。

②范为:《大数据时代个人信息保护的路径重构》,《环球法律评论》2016年第5期。

个人信息,也包含着对这些信息运用场景的限制,网站对个人信息的分析、使用方式,应当以网站功能、访问者浏览内容为限。例如访问者将个人的健康生理信息提交至医院网站,目的是让医疗人员了解自身的健康状况,从而有针对性地提供医疗咨询等服务,或是由医院对多名患者的个人信息进行去标识化处理,使它们不再具备识别特定自然人身份的功能,并在此前提下进行大数据分析,以此为依据改进诊疗方案。网络爬虫的使用者即使收集了与网站管理者相同范围的个人信息,也由于没有明确告知访问者收集信息的事实,以及收集信息之后的用途,超出了访问者对个人信息使用场景的预期,使得这些信息在泄露之后产生不可预知的风险,因此前述收集信息的行为具有非法性。

例如在“沈某侵犯公民个人信息案”中,被告人沈某为获取利益,于 2018 年创建了一个名为“医院竞价管理系统”的公民信息抓取系统,该系统通过在医院网站内写入网络爬虫代码的方式,利用移动网络漏洞接口,在网民使用手机访问互联网网站时,抓取“病患”等关键词,同时获取该网民上网时对应的手机号码、访问 IP、手机归属地等信息共计 491 040 条,并将获取的这些信息存入数据库,销售给需要推广业务的医院。法院认定沈某的行为属于非法获取并出售公民个人信息,因为无证据证明这些信息是沈某在网民授权或同意的情况下获取的。但由于这些个人信息并非关键、重大的人身、财产、行为信息,因此可在量刑时予以考虑①。

此外,该情形下还需要对外部访问者的信息类型予以甄别。相较于《中华人民共和国网络安全法》仅将“可识别性”确立为公民个人信息的判断标准,《个人信息刑案解释》指出,刑法意义上的“公民个人信息”应当具备识别身份、反映活动情况这两项功能之一,这体现了公民个人信息的范围扩张。从前述两项功能的原理而言,“识别身份”是从信息出发,确定特定个体是否符合预设的识别条件;“反映活动情况”则是从特定个体出发,对基于其活动记录而产生的信息予以组合,分析这些信息是否具有反映活动情况的功能②。但在甄别访问者浏览网页所产生的信息时,面临着“数位足迹”能否被认定为刑法意义上“公民个人信息”的问题。“数位足迹”是指特定个体在从事网络活动过程中所产生的信息与资料,具体体现为三种类型:第一,与网页浏览相关的记录,主要包括访问者输入的搜索关键词、所浏览的商品或服务类型、在网站的各个区域所浏览的时长。第二,与登录信息相关的记录,主要包括能够在一段时间内提供免登录功能的身份认证信息,具体由访问者在网站所注册的账户名称与密码组成。第三,与访问者的技术条件相关的记录,主要包括其所使用的设备类型、IP 地址、网络接入服务类型。

这些信息虽然反映了特定个体在访问网站过程中的行为偏好,但不应被认定为刑法意义上的“公民个人信息”。一方面,这些信息与特定个体之间缺乏直接的关联,例如 IP 地址所对应的是计算机等电子设备,访问者所使用的账户名称与密码只对应虚拟的个体,其中未包含能够识别访问者身份的信息。一般人在未采取特殊技术手段的情况下,也无法将这些信息与其他信息相结合,用于识别特定个体的身份。另一方面,不应将访问者在网络上留下的活动轨迹、信息痕迹,与《个人信息刑案解释》中的“行踪轨迹信息”相混同。行踪轨迹信息属于《个人信息刑案解释》中的高度敏感信息,适用最严格的保护标准,因此在外延上应当限制与缩小。有关的指导案例指出,当公民从事某些活动不希望被他人获悉时,因其行踪轨迹与所从事的活动之间具有直接联系,一旦被他人获悉,其所从事的活动也就相当程度地被暴露,此时行踪轨迹由于具有明显的隐私性和权益性,因此属于刑法所保护的公民个人信息③。具体而言,行踪轨迹信息应当具备直接定位特定自然人具体坐标的功能,例如通过 GPS 设备、车辆定位系统所反映的信息。而诸如火车票、网约车行程单上的起始地点与启程时间信息,则通常不宜纳入行踪轨迹信息的范围④。因此,由于网络空间

①西安市未央区人民法院(2019)陕 0112 刑初 880 号刑事判决书。

②杨楠:《个人数位足迹刑法规制的功能性偏误与修正》,《安徽大学学报(哲学社会科学版)》2019 年第 4 期。

③罗灿,李永京,徐辉:《谢新冲出售公民个人信息案——手机定位属于刑法保护的“公民个人信息”》,《刑事审判参考》2011 年第 6 期。

④喻海松:《侵犯公民个人信息罪的司法适用态势与争议焦点探析》,《法律适用》2018 年第 7 期。

中的访问记录不同于现实社会中的行踪轨迹,而且不能直接定位访问者所处的具体坐标,故不能反映自然人的活动情况,不属于个人信息的范畴。但由于访问者在与网站进行信息交互的过程中,活动轨迹、信息痕迹属于《中华人民共和国刑法》第 285 条中的"计算机信息系统中传输的数据",因此可对相关的非法获取行为适用非法获取计算机信息系统数据罪予以规制。

结语

数字经济时代对网络爬虫技术的运用,既需要充分发挥其高效收集信息的优势,又需要兼顾个人信息权益的法律保护,使工具理性受到价值理性的约束,在规范网络爬虫使用的过程中以法律为主体、以技术为补充。面对数据权利化与数据分享性的理念冲突,首先应厘清数据与信息之间的关系,划分刑事惩治与其他部门法规制措施之间的界限,使得网络爬虫技术的入罪边界、责任类型得以明确;其次应类型化、场景化地分析相关的刑法罪名适用问题;最后应在法律适用过程中限定个人信息的外延,充分平衡网络爬虫技术使用者、个人信息存储主体、个人信息对应的公民等各方的利益,构建个人信息利用与保护的均衡机制。

Criminal Law Regulation of Abusing Web Crawler Technology to Collect Personal Information

SONG Xing-jian

(School of Criminal Justice, China University of Political Science and Law, Beijing 100088, China)

Abstract: With the continuous improvement of information collection ability of web crawler technology, the loss caused by its abuse is increasingly serious. To identify "illegal collection" of personal information, we should examine not only whether the method itself is illegal, but also the basis or qualification of information collection. When breaking through the anti-crawler technical measures to collect personal information, according to the concurrence relation between articles of criminal law, the crime of infringing on citizens' personal information should be applied to. When logging into the system normally to collect personal information, we should check whether the collection behavior is necessary for the performance of duties. When collecting the personal information of the external visitors of the website, the authority of information collection should be determined according to the website function and browsing content.

Key words: web crawler; personal information; electronic data; illegal collection; minimum necessity

(责任校对　龙四清)