

数据爬取的 合理边界与法律风险

上海市锦天城律师事务所

吴卫明 博士|高级合伙人

2019年去10月11日



上海

吴卫明

产业经济学
电子商务 博士 **律师**

上海市锦天城律师事务所高级合伙人
网络金融、资产管理、电子商务法律专家
上海律师协会互联网业务委员会副主任
多所知名大学法学院兼职研究生导师
2016年度legalband 中国十佳互联网金融律师
2018年度legalband 中国十佳网络安全与数据保护律师
2019年ALB(《亚洲法律杂志》)中国首届十五佳TMT行业律师
2019年度legalband 中国十佳金融科技律师



目录

CATALOG



数据爬取简介



数据爬取的法律风险



进行数据爬取的合理边界

上海“数据治理系列沙龙”第四期

01



数据爬取简介

上海“数据治理系列沙龙”第四期

概念

类型

对象

用途



概念

数据爬取，是指利用“爬虫”在互联网上抓取信息（数据）的行为或过程。

爬虫，也叫网络爬虫、网页蜘蛛、网络机器人（robots），是一种按照一定的规则，自动地抓取互联网信息的程序或者脚本，它可以系统地搜索互联网并下载网页副本，然后通过搜索引擎对其进行索引。

在当前的网络环境下，爬虫爬取数据的对象还包括在获得特定信息系统或网络空间权利人同意后对相关信息系统或空间内有关数据的爬取。如同业爬虫等。

爬虫

数据
爬取

数据



概念



数据收集与数据爬取，可以说数据爬取是数据收集的一种方式，主要依赖于自动化的程序（脚本），是一种高效的数据收集方式。



类型

通用爬虫

聚焦爬虫

通用爬虫：是搜索引擎抓取系统（Baidu、Google、Yahoo等）的重要组成部分。主要目的是将互联网上的网页下载到本地，形成一个互联网内容的镜像备份。

聚焦爬虫：是“面向特定主题需求”的一种网络爬虫程序，它与通用搜索引擎爬虫的区别在于：聚焦爬虫在实施网页抓取时会对内容进行处理筛选，尽量保证只抓取与需求相关的网页信息。



对象

数据

按是否筛选

按是否公开

按数据内容

网页整体信息

筛选后的信息

公开数据

非公开数据

个人数据

作品数据

其他数据

数据爬取的对象是网页上或特定信息系统内的数据，而数据本质上是一种具有特定表现形式的信息。同时，数据本身在文字意义上、法律意义上都可能有复杂的类型、属性。爬取不同类型的数据会有不同的法律要求和法律后果。



用途

搜索引擎

通用网络爬虫 从互联网中搜集网页，采集信息，这些网页信息用于为搜索引擎建立索引从而提供支持，它决定着整个引擎系统的内容是否丰富，信息是否即时。

数据分析

聚焦爬虫在实施网页抓取时会对内容进行处理筛选，尽量保证只抓取与需求相关的网页信息。根据数据分析的不同需求，可以通过聚焦爬虫抓取到特定字段的信息。数据分析可以应用于行业分析、趋势判断、风险控制等领域。

精准营销

在聚焦爬虫中，基于联系方式的爬取，如电话、微信、QQ、电子邮箱，结合通过其他爬取的信息进行数据分析，可以实现精准营销的前期准备。

02



数据爬取的法律
风险

上海“数据治理系列沙龙”第四期

数据的法律属性

非法爬取不同数据对应的法律风险



数据的法律属性

个人隐私

用户在网上可以通过设置访问限制、加密功能对属于自己隐私的数据进行保护。

01



02



商业秘密

与个人隐私相似，虽然网络上公开的信息较难构成商业秘密，但由于网络上某些信息可以通过采取技术措施使得仅有特定的用户可以接触，仍有可能具备秘密性和保密性，构成商业秘密。

03



商业资源

数据本身可能需要经营者花费一定成本得到，并可以为经营者带来商业利益，可以认为是一种应当保护的商业资源。

个人信息

即使对于不属于个人隐私范畴的个人信息，法律仍然拥有一定的保护规定。

06



05



作品

数据可以具有多种形式，具备构成文字作品、图片作品、影像作品的可能性，可以通过著作权加以保护。

04



信息系统内容物

数据是保存在特定载体上的，可以是信息系统的内容，而法律对于信息系统有相应保护规定。



非法爬取不同数据对应的法律风险

数据的法律属性	个人隐私/个人信息	商业秘密	其他商业资源	作品	信息系统内容物
是否可能承担侵权责任	√	√	√	√	√
是否可能承担行政责任	√	√	√	√	√
是否可能承担刑事责任	√	√	×	× (仅获取不构成犯罪, 发行、销售可能构成)	√ (使用侵入方式、为侵入提供工具)

上海“数据治理系列沙龙”第四期



关于个人隐私与个人信息

民法总则

- 第一百一十条第一款 自然人享有生命权、身体权、健康权、姓名权、肖像权、名誉权、荣誉权、**隐私权**、婚姻自主权等权利。
- 第一百一十一条 自然人的**个人信息**受法律保护。任何组织和个人需要获取他人个人信息的，应当依法取得并确保信息安全，不得非法收集、使用、加工、传输他人个人信息，不得非法买卖、提供或者公开他人个人信息。

最高人民法院关于审理利用信息网络侵害人身权益 民事纠纷案件适用法律若干问题的规定

- 第十二条 网络用户或者网络服务提供者利用网络公开自然人基因信息、病历资料、健康检查资料、犯罪记录、家庭住址、私人活动等个人隐私和其他个人信息，造成他人损害，被侵权人请求其承担侵权责任的，人民法院应予支持。但下列情形除外：……

消费者权益保护法

- 第二十九条 经营者收集、使用消费者个人信息，应当遵循合法、正当、必要的原则，明示收集、使用信息的目的、方式和范围，并经消费者同意。经营者收集、使用消费者个人信息，应当公开其收集、使用规则，不得违反法律、法规的规定和双方的约定收集、使用信息。
经营者及其工作人员对收集的消费者个人信息必须严格保密，不得泄露、出售或者非法向他人提供。经营者应当采取技术措施和其他必要措施，确保信息安全，防止消费者个人信息泄露、丢失。在发生或者可能发生信息泄露、丢失的情况下，应当立即采取补救措施。
经营者未经消费者同意或者请求，或者消费者明确表示拒绝的，不得向其发送商业性信息。

上海“数据治理”沙龙”第四期

关于个人隐私与个人信息

网络安全法

- 第二十二条第三款 网络产品、服务具有收集用户信息功能的，其提供者应当向用户明示并取得同意；涉及用户个人信息的，还应当遵守本法和有关法律、行政法规关于个人信息保护的规定。
- 第四十一条 网络运营者收集、使用个人信息，应当遵循合法、正当、必要的原则，公开收集、使用规则，明示收集、使用信息的目的、方式和范围，并经被收集者同意。网络运营者不得收集与其提供的服务无关的个人信息，不得违反法律、行政法规的规定和双方的约定收集、使用个人信息，并应当依照法律、行政法规的规定和与用户的约定，处理其保存的个人信息。
- 第四十二条 网络运营者不得泄露、篡改、毁损其收集的个人信息；未经被收集者同意，不得向他人提供个人信息。但是，经过处理无法识别特定个人且不能复原的除外。网络运营者应当采取技术措施和其他必要措施，确保其收集的个人信息安全，防止信息泄露、毁损、丢失。在发生或者可能发生个人信息泄露、毁损、丢失的情况时，应当立即采取补救措施，按照规定及时告知用户并向有关主管部门报告。
- 第四十三条 个人发现网络运营者违反法律、行政法规的规定或者双方的约定收集、使用其个人信息的，有权要求网络运营者删除其个人信息；发现网络运营者收集、存储的其个人信息有错误的，有权要求网络运营者予以更正。网络运营者应当采取措施予以删除或者更正。
- 第四十四条 任何个人和组织不得窃取或者以其他非法方式获取个人信息，不得非法出售或者非法向他人提供个人信息。

全国人民代表大会常务委员会关于加强网络信息保护的决定

- 二、网络服务提供者和其他企业事业单位在提供服务活动中收集、使用公民个人电子信息，应当遵循合法、正当、必要的原则，明示收集、使用信息的目的、方式和范围，并经被收集者同意，不得违反法律、法规的规定和双方的约定收集、使用信息。网络服务提供者和其他企业事业单位收集、使用公民个人电子信息，应当公开其收集、使用规则。

关于个人隐私与个人信息

刑法

- 第二百五十三条之一【**侵犯公民个人信息罪**】违反国家有关规定，向他人出售或者提供公民个人信息，情节严重的，处三年以下有期徒刑或者拘役，并处或者单处罚金；情节特别严重的，处三年以上七年以下有期徒刑，并处罚金。

违反国家有关规定，将在履行职责或者提供服务过程中获得的公民个人信息，出售或者提供给他人的，依照前款的规定从重处罚。

窃取或者以其他方法非法获取公民个人信息的，依照第一款的规定处罚。

单位犯前三款罪的，对单位判处罚金，并对其直接负责的主管人员和其他直接责任人员，依照各该款的规定处罚。



余某侵犯公民个人信息案

主要事实

- 余某在某宝任职期间使用python语言编写具有自动获取数据功能的脚本程序main_user.py，并将其在内部论坛网站账号的cookie信息配置到该脚本程序上，通过运行该程序，秘密取得该集团员工的个人信息共计2万余条
- 余某离职时，将上述信息存储于电脑硬盘秘密带走

裁判要点

- 余某的权限仅为查询、浏览，为了获取整体组织架构而积极、主动编写获取信息的脚本，在无人授权且知晓的情况下，于二个多月的任职期内多次使用脚本爬取员工个人信息数据并保存，随后离职将上述数据偷偷带走，属于违法的“窃取”行为
- 被窃取的员工个人信息内容包括姓名、工号、部门、支付宝账号、职位、昵称、生日、性别、工作电话、邮箱等30种以上，属于“公民个人信息”
- 被窃取的员工个人信息数量多达2万余条，且其中均包含支付宝账号等财产信息，属于“情节严重”

裁判结果

- 构成非法获取公民个人信息罪

关于商业秘密与不正当竞争

反不正当竞争法

• 第二条第二款 本法所称的不正当竞争行为，是指经营者在生产经营活动中，违反本法规定，扰乱市场竞争秩序，损害其他经营者或者消费者的合法权益的行为。

• 第九条 经营者不得实施下列侵犯商业秘密的行为：

- (一) 以盗窃、贿赂、欺诈、胁迫、电子侵入或者其他不正当手段获取权利人的商业秘密；
- (二) 披露、使用或者允许他人使用以前项手段获取的权利人的商业秘密；
- (三) 违反保密义务或者违反权利人有关保守商业秘密的要求，披露、使用或者允许他人使用其所掌握的商业秘密；
- (四) 教唆、引诱、帮助他人违反保密义务或者违反权利人有关保守商业秘密的要求，获取、披露、使用或者允许他人使用权利人的商业秘密。

经营者以外的其他自然人、法人和非法人组织实施前款所列违法行为的，视为侵犯商业秘密。

第三人明知或者应知商业秘密权利人的员工、前员工或者其他单位、个人实施本条第一款所列违法行为，仍获取、披露、使用或者允许他人使用该商业秘密的，视为侵犯商业秘密。

本法所称的商业秘密，是指不为公众所知悉、具有商业价值并经权利人采取相应保密措施的技术信息、经营信息等商业信息。

刑法

• 第二百一十九条【侵犯商业秘密罪】有下列侵犯商业秘密行为之一，给商业秘密的权利人造成重大损失的，处三年以下有期徒刑或者拘役，并处或者单处罚金；造成特别严重后果的，处三年以上七年以下有期徒刑，并处罚金：

- (一) 以盗窃、利诱、胁迫或者其他不正当手段获取权利人的商业秘密的；
- (二) 披露、使用或者允许他人使用以前项手段获取的权利人的商业秘密的；
- (三) 违反约定或者违反权利人有关保守商业秘密的要求，披露、使用或者允许他人使用其所掌握的商业秘密的。

明知或者应知前款所列行为，获取、使用或者披露他人的商业秘密的，以侵犯商业秘密论。

本条所称商业秘密，是指不为公众所知悉，能为权利人带来经济利益，具有实用性并经权利人采取保密措施的技术信息和经营信息。

本条所称权利人，是指商业秘密的所有人和经商业秘密所有人许可的商业秘密使用人。



央视诉视畅公司不正当竞争案

主要事实

- 央视经国际奥委会授权，在中国境内独家享有通过信息网络，转播中央电视台制作、播出伦敦奥运会开幕式电视节目的权利
- 视畅公司未经许可，通过其运营的www.kanketv.com网站及由其发布的“看客影视”安卓系统客户端软件，向公众实时转播伦敦奥运会开幕式节目，视畅公司采用的方式是利用网络爬虫技术获取了源于www.cntv.cn网站中的涉案视频链接并向社会公众提供
- 央视于www.cntv.cn网站发布《关于独家享有伦敦2012年奥运会新媒体版权的声明》

裁判要点

- 视畅公司提供开幕式节目的搜索链接，因其指向影像源自央视网站，故不构成著作权侵权
- 视畅公司提供搜索链接的行为构成不正当竞争——在未偿付对价的情况下，利用央视保留权利的商业资源牟利并致央视利益受损（因浏览页面未发生变化，故相关公众根据其感官判断，会认为所观赏的开幕式节目是由www.kanketv.com网站提供）

裁判结果

- 视畅公司赔偿央视经济损失及合理开支



大众点评网诉百度不正当竞争案

主要事实

- 大众点评网收集了大量商户信息，并吸引大量消费者通过真实体验发布点评信息。大众点评网中的用户点评等内容已经成为广大消费者选择相关商家和服务的重要参考资料，取得了良好的社会效益和经济效益
- 百度爬取大众点评网的点评信息未违反大众点评网上的robots协议
- 百度未经许可，在百度地图、百度知道中大量抄袭、复制大众点评网的点评信息，直接替代了大众点评网向用户提供内容

裁判要点

- 爬取大众点评网信息的行为不构成不正当竞争
- 节选点评信息部分内容并提供链接的行为不构成不正当竞争
- 展示完整点评信息的行为构成不正当竞争

裁判结果

- 百度立即停止以不正当的方式使用大众点评网的点评信息
- 百度赔偿大众点评网经济损失及为制止不正当竞争行为所支付的合理费用



大众点评网诉爱帮网的著作权侵权纠纷案

“数据治理系列”第四期

主要事实

- 大众点评网中的商户介绍和用户点评已经成为广大消费者选择相关商家和服务的重要参考资料，并进而为大众点评网取得了良好的社会效益和经济效益。大众点评网投入大量人力和资金创作了相关商户介绍信息，并享有其全部著作人身权和财产权。根据大众点评网与大众点评网注册用户签订的在线协议，大众点评网独家拥有大众点评网中所载用户点评内容的著作财产权
- 爱帮网公司未经原告同意，擅自复制原告享有著作权的商户介绍和点评，刊登于其经营管理的爱帮网上，试图建立“最大、最全的生活信息网上平台”，并在全国范围内进行广告招商活动

裁判要点

- 大众点评网未能证明享有涉案商户介绍的著作权
- 涉案点评内容主要是消费者的主观感受，体现了作者独特的思想、个性、情感、体验，在表现形式体现了一定的风格和特点，具有一定的独创性，依法构成作品
- 大众点评网拥有点评内容的相应著作权利
- 垂直搜索引擎技术不违法
- 使用大众点评网享有著作权的点评构成侵权

裁判结果

- 停止在爱帮网（www.aibang.com）使用大众点评网享有著作权的作品
- 爱帮网赔偿大众点评网经济损失及诉讼合理开支



关于信息系统

网络安全法

刑法

【非法获取计算机信息系统数据、非法控制计算机信息系统罪】违反国家规定，侵入前款规定以外的计算机信息系统或者采用其他技术手段，获取该计算机信息系统中存储、处理或者传输的数据，或者对该计算机信息系统实施非法控制，情节严重的，处三年以下有期徒刑或者拘役，并处或者单处罚金；情节特别严重的，处三年以上七年以下有期徒刑，并处罚金。

【提供侵入、非法控制计算机信息系统程序、工具罪】提供专门用于侵入、非法控制计算机信息系统的程序、工具，或者明知他人实施侵入、非法控制计算机信息系统的违法犯罪行为而为其提供程序、工具，情节严重的，依照前款的规定处罚。

单位犯前三款罪的，对单位处罚金，并对其直接负责的主管人员和其他直接责任人员，依照各该款的规定处罚。

03



进行数据爬取的
合理边界

上海“数据治理系列沙龙”第四期

对象合法

手段合法

权利人同意



对象合法

判断爬取
的数据类
型

避免爬取
禁止收集
的数据类
型

根据数据
类型设置
不同的准
入门槛

“数据合规”系列沙龙 第四期



准入门槛——以个人信息为例

根据需求判断爬取的数据必要性

避免爬取超过必要范围的数据类型

根据数据类型判断敏感度

为敏感个人信息类型设置更严格的获取同意的方式

上海“数据治理系列沙龙”第四期



手段合法





手段合法

《数据安全管理办法（征求意见稿）》

- 第十六条 网络运营者采取自动化手段访问收集网站数据，不得妨碍网站正常运行；此类行为严重影响网站运行，如自动化访问收集流量超过网站日均流量三分之一，网站要求停止自动化访问收集时，应当停止。



对于刑事非法性的思考

《刑法》第253条规定“窃取或者以其他方法非法获取公民个人信息的，依照第一款的规定处罚。”

司法解释第四条规定：违反国家有关规定，通过购买、收受、交换等方式获取公民个人信息，或者在履行职责、提供服务过程中收集公民个人信息的，属于第二百五十三条之一第三款规的“以其他方法非法获取公民个人信息”。

“违反国家有关规定”，成为判断的基础。基于立法法关于设定刑罚的权限，有关规定应属于“法律”层级。

《网络安全法》第四十一条：网络运营者收集、使用个人信息，应当遵循合法、正当、必要的原则，公开收集、使用规则，明示收集、使用信息的目的、方式和范围，并经被收集者同意。

网络运营者不得收集与其提供的服务无关的个人信息，不得违反法律、行政法规的规定和双方的约定收集、使用个人信息，并应当依照法律、行政法规的规定和与用户的约定，处理其保存的个人信息。

《民法总则》第一百一十一条：自然人的个人信息受法律保护。任何组织和个人需要获取他人个人信息的，应当依法取得并确保信息安全，不得**非法收集、使用、加工、传输**他人个人信息，不得非法买卖、提供或者公开他人个人信息。

对于爬取第三方信息系统内的数据问题，有两方面的考量：

- 1、未经特定空间权利人许可的数据爬取：可能构成侵入计算机信息系统，从而涉嫌非法获取计算机信息系统数据罪；
- 2、对于经过特定权利人许可（表现为提供账户名和密码）的数据爬取，则需要区分行为违法性的范畴。避免对侵犯公民个人信息刑事违法性判断的不当扩大。

公开不足或超过必要性
引发的民事纠纷？

超过必要性
引发的行政处罚？

以其他方法非法获取
公民个人信息？
(侵犯公民个人信息)

公开收集、使用规则，明示收集、使用信息的目的、方式和范围，并经被收集者同意

不得违反法律、行政法规的规定和双方的约定收集个人信息

不得收集与其提供的服务无关的个人信息

不得违反法律、行政法规的规定和双方的约定使用个人信息

并应当依照法律、行政法规的规定和与用户的约定，处理其保存的个人信息

向他人出售或者
提供公民个人信息
(侵犯公民个人信息)

《网络安全法》第四十一条



权利人同意

著作
权人

针对属于作品的数
据，爬取前应得到著
作权人的同意

个人信
息主
体

针对属于个人隐私或个
人信息的数据，爬取
前应得到个人信息主
体的同意

网站
运营者

针对属于特定网站上的数
据，爬取前建议得
到网站运营者的同意，
但可以证明网站运营
者公开同意或无权禁
止的除外

信息系
统权
利人

针对属于特定信息系
统上的数据，爬取
建议通过信息系统权
利人同意的方式进行

其他
权利人

如果有证据表明存在
对数据享有权利其
他主体，建议爬取前
取得相应主体同意





THANKS

上海“数据治理沙龙”第四期