数据爬取问题的实务应对

顾伟 博士

内容仅代表个人研究观点

上海“数据治理系列沙龙”第四期

Copyright © 2019 by Alibaba Group.
阿里巴巴集团版权所有

This material is Confidential.
本文档所有信息均为保密信息

This material is for internal use only.
本文档仅限于内部交流使用

上海“数据治理系列沙龙”第四期

1. 数据爬取的法律风险分析

国内

刑事

★★★☆☆

涉多重罪名

非法获取计算机信息系统数据罪

侵犯著作权罪

侵犯公民个人信息罪

民事

★★★★☆

不正当竞争

以不正当竞争案件为主

部分涉及侵犯信息网络传播权案

行政

★☆☆☆☆

执法很少

证据认定等困难，行政执法少

上海“数据治理系列沙龙”第四期

1. 数据爬取的法律风险分析

国内

淘宝诉美景不正当竞争案，法院判决重点关注：

1、美景的经营行为，是否违反诚信原则与商业道德（不劳而获、搭便车）

2、淘宝对数据产品是否具有法定权益
- 劳动付出、经营积累，带来商业利益、市场竞争优势
- 原生数据，无独立权利；衍生数据，独立财产权益

宋健候明强等非法获取计算机信息系统数据案：
被告使用device.ID 绕过服务器身份校验，使用伪造UA及IP绕过服务器访问频率限制爬取数据

邵凌霜、陈昂等非法获取计算机信息系统数据案：
被告破解“酷米客”客户端的加密算法，使用变化的IP绕过服务器访问频率限制爬取数据

上海“数据法理系列沙龙”第四期

1. 数据爬取的法律风险分析

美国

主要法律依据：《1986年计算机欺诈与滥用法》CFAA

未经授权故意访问计算机或超过授权访问权限，从而从任何受保护的计算机获取信息；或者被告“故意造成程序传输，并且对未经授权且受保护的计算机造成损害”

根据判例，法院们对“授权”的理解并不一致，但有对爬虫宽容的趋势：

1、传统的，网站有明确的声明、协议、告知、警告、弹窗禁止爬虫可能被法院认定为CFAA规定的“未经授权”或“超出授权”的行为，依此可以予以确立刑事、民事责任。

2、越来越多法院不认可事前声明等方式，但是事后的专门通知和技术手段依然被法官予以认可。

上海“数据治理系列沙龙”第四期

1. 数据爬取的法律风险分析

美国

2019 年 9 月 9 日，美国联邦第九巡回法院作出上诉判决:维持加州北区联 邦法院就 hiQ Labs 诉 LinkedIn 案颁布的诉中禁令。

为避免与第一修正案冲突，法院对“授权”进行了狭义解释，“未经授权访 问”这一表达本身，喻示相应的“受保护的计算机”并非任由公众访问。因此，CFAA只保护受密码保护网站或者，其它限制一般公众浏览之上信息的网站

CFAA 此处禁止的是类似“破坏并闯入(break and enter)”计算机的行为，而非仅仅“使用(use)”计算机的行为。 CFAA 是一部“反侵入(anti-intrusion)”法，而非反“滥用 (misappropriation)”法

上海“数据治理系列沙龙”第四期

2. 数据爬取问题的实务困境

越来越绕不过的刑事责任

- 开放的PC空间落幕 半封闭的APP成为趋势
- APP爬取绕不过对加密技术的破解

越来越肆无忌惮的竞争

- 搭便车的行为风起
- 竞争损害越发严重

越来越需要开放的数据

- AI技术的发展 数据需求无止境
- 有序开放能够最大化社会福祉

上海“数据治理系列沙龙”第四期

3. 关于数据爬取问题的思考

坚持刑法的谦抑性

着重通过反不正当竞争法与著作权法解决争议

立法规范爬取手段确保利益平衡

推动网络平台有序开放数据

包容审慎

上海“数据治理系列沙龙”第四期

4. 对相关网络运营者的建议

1 协议保障

明确主张数据权属，禁止爬虫

2 技术措施

APP接口加密、访问IP限制等

3 反爬取证

对明显侵害合法权益的爬取行为进行取证

4 商业合作

优先考虑，积极寻求商业合作空间

5 举报、起诉

明显违法或者侵害公司合法权益的，可结合现有国内外案例，进行举报或起诉

上海“数据治理系列沙龙”第四期

上海“数据治理系列沙龙”第四期

谢谢!

Confidential-For Internal Use Only