

数据爬虫合法边界 研究

腾讯网络安全与犯罪研究基地 高级研究员 张宝峰

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



网络爬虫技术的应用

搜索引擎



数据聚合

数据挖掘



网络安全

上海“数据治理系列沙龙”第四期

君子协定 ——Robots协议

```
User-agent: Baiduspider  
Disallow: /baidu  
Disallow: /s?  
Disallow: /ulink?  
Disallow: /link?  
Disallow: /home/news/data/
```

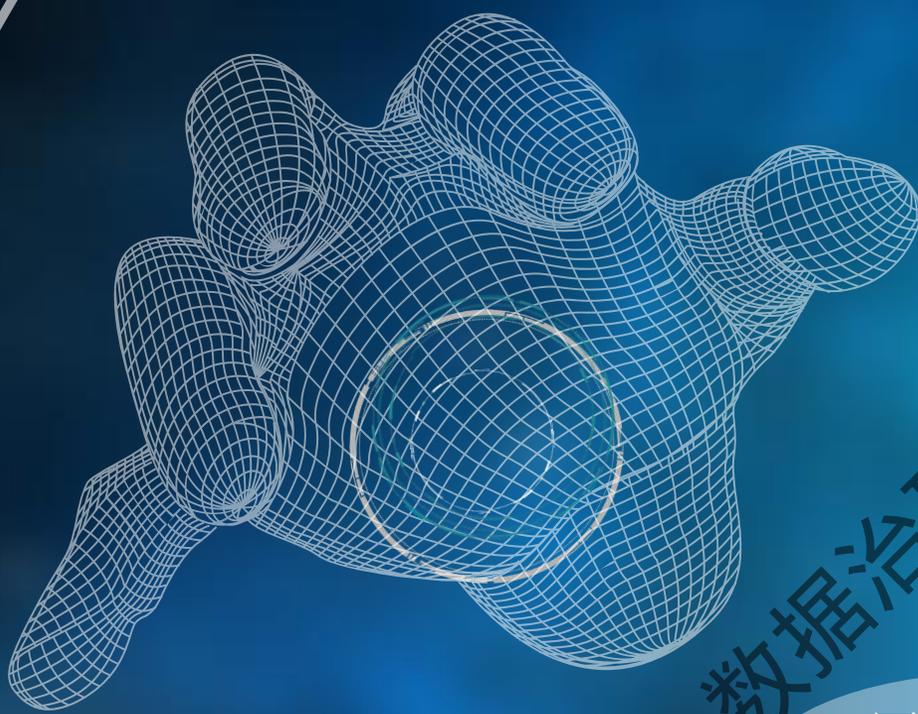
```
User-agent: Googlebot  
Disallow: /baidu  
Disallow: /s?  
Disallow: /shifen/  
Disallow: /homepage/  
Disallow: /cpro  
Disallow: /ulink?  
Disallow: /link?  
Disallow: /home/news/data/
```



“数据治理系列沙龙”第四期

国外有关网络爬虫技术典型案例

时间	案件事实
2003	EFCultural Travel BV v. ZeferCorp., 318 F.3d 58, 62
2004	SouthwestAirlines Co. v. Farechase,Inc., 318 F. Supp.2d 435, 439 – 440
2013	CraigslistInc. v. 3Taps Inc., 942 F. Supp. 2d 962
2016	EarthCam,Inc. v. OxBlueCorp., 703 Fed.Appx.803, 808
2019	hiQLabs, Inc. v. LinkedIn Corp., 273 F. Supp.3d 1099



我国有关网络
爬虫的法律规
制及相关案例

民事审判
案例

刑事审判
案例

上海“数据治理系列沙龙”第四期

恶意爬虫的防止措施

采取反爬技术措施



设置版权信息



设置明确的反爬虫声明



设置版权信息



及时监控，积极维权

上海“数据治理系列沙龙”第四期

网络爬虫不当使用涉嫌侵权犯罪的情景

- 1.爬取非公开信息；
2. 爬取非公开且未经用户授权的公民个人信息；
3. 爬取虽属公开信息但他人享有著作权的作品；
4. 爬取虽属公开信息但双方存在直接竞争关系，构成作品实质性替代的；
5. 网络爬虫技术获取的信息即使属于公开信息且经授权，非法提供仍可能构成违法犯罪；
6. 即使经过用户授权，爬取是否合法，也存在一定争议。

网络爬虫技术的合理应用

遵守Robots协议；

检查反爬虫措施

限制数据使用目的



区别爬取数据的性质

控制爬取频率和数量

谢谢!



腾讯网络安全与犯罪研究基地

上海“数字法治”第四期