

**实 习 周 报**

**作者：林晓丹**

**日期：2017.07.15**

**广州探迹科技有限公司 版权所有**

1 本周工作

|  |  |
| --- | --- |
| 任务项 | 完成进度 |
| TY\_0001：高德逆地理编码API爬取 | 100% |
| TGNB\_PARSE\_12：腾讯证券港股信息 | 100% |
| TGNB\_PARSE\_13：天眼查企业全称爬取 | 100% |
| GW0016：域名页面下所有URL获取 | 100% |
| GW0017：读取域名查询IP | 100% |

# 2 学习计划

|  |  |
| --- | --- |
| 计划项 | 进度/计划时间 |
| js | 70%/四周 |

# 3 心得体会

本周完成了三个咨询组的爬虫需求，都比较简单。之后就在有需求时启动爬虫，监控爬虫的结果。周一整理了一份数据到TY\_0001的爬虫上跑，发现在整合数据的时候比较容易有坑，需要大致清楚数据的格式再提取，不然就很容易因为报错或者格式问题忽略掉部分数据。

周五完成了两个域名的爬虫，主要是在域名的提取和处理上花费比较多的时间，需要考虑到不同情况的URL需要调整的格式，同时了解了urllib的一些方法。