2010年到2013年期间，我们主要为澳大利亚(Webscraping)和美国(Vitals)公司提供网站数据采集外包服务，先后成功完成了超过600个项目，涉及网站上千个。这是我们一个重要的成长阶段，为公司累积下来了丰富的Web分析和数据提取实战经验，尤其是在应对各种网站反采集策略上。下面简单介绍一些这些项目：

根据项目内容，大致可以分为如下三种类型：

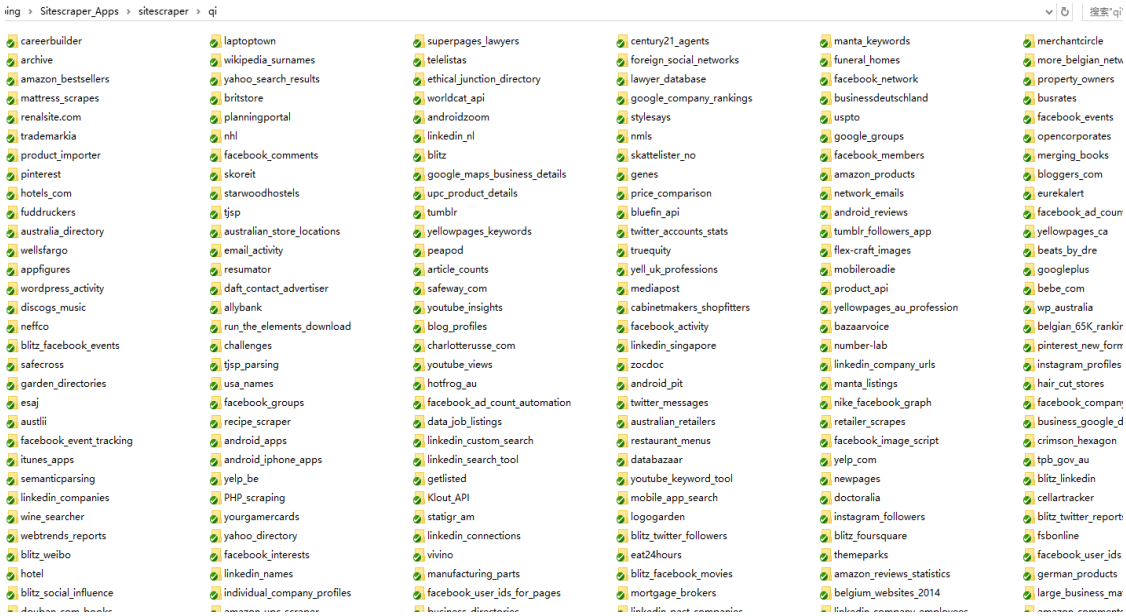
1. 数据定制抓取。客户最终需要的只是数据。约占70%。
2. 采集器（脚本）定制。客户最终需要的是采集脚本。约占25%。
3. Web API开发。客户最终需要的是一个Web接口。约占5%。

根据项目难度大致可以分为如下四种类型：

1. 简单。网站无防护策略，内容静态展示。约占30%。
2. 中等。网站无防护策略，内容动态加载，例如Ajax技术。约占40%。
3. 较难。网站有封IP策略。约占20%。
4. 非常难。网站封IP很厉害、数据有加密。约占10%。

这些采集项目的开发语言均为Python，运行平台80%为Linux平台（Ubuntu为主）、20%为Windows平台。数据输出格式80%为CSV，10%为JSON和XML，10%为MySQL和SQL Server。

附：部分项目列表。



公司名称：西安鲲之鹏网络信息技术有限公司

网　　址：http://www.site-digger.com/

地　　址：陕西省西安市雁塔区长安中路99号长安文化综合大厦11821室

邮　　编：710061

联系电话：029 - 87553281

手　　机：13571845363 齐工

13389148466 周工
客 服 QQ： 1649677458 或 312602670
邮　　箱： hello@site-digger.com