

甲方名称:

上海某大数据公司（因保密协议限制无法公开具体信息）

甲方需求:

根据需求定制一采集平台，并进行长期维护。

1. 采集目标: 百度指数系统, <http://index.baidu.com>;
2. 从数据库读取待查询关键词, 依次在百度指数系统中进行查询;
3. 提取查询结果的日期和百度指数值存储到数据库中;
4. 采集频率: 每日采集一次, 保留每次采集结果;
5. 采集量: 查询关键词约2万个;

技术难点:

- 百度指数系统需要登录才能使用;
- 百度指数系统封IP很厉害, 容易出现验证码;

实现方案:

- 使用多个账号自动登录采集;
- 使用独享稳定高匿HTTP代理绕过封IP限制;
- 通过验证码处理技术解决验证码问题;

项目状态:

本项目于2016年6月1日正式开始上线运行, 目前仍在稳定进行中。



西安鲲之鹏网络信息技术有限公司

选择我们, 所有数据都是你的!



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