任务名称： 依据域名清单爬取站长工具的IP地址及归属地

任务编号： ALY\_CRAWL\_01

1. 具体步骤：
2. 访问如下链接：

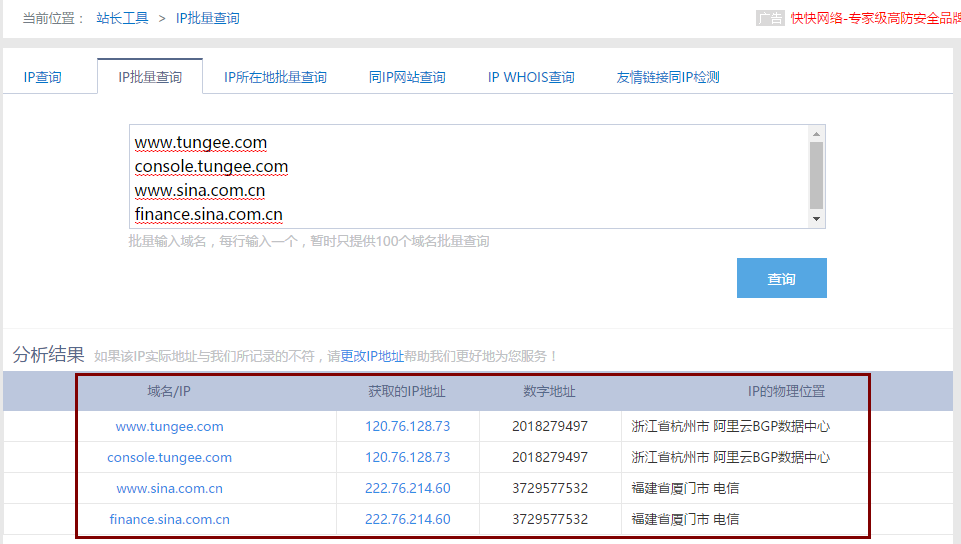
http://ip.chinaz.com/ipbatch



1. 输入域名列表，一次最多支持100个，点击“查询”按钮，参见下图：



1. 爬取IP与归属地结果：



4. 解析维度：

|  |  |
| --- | --- |
| 原始域名 | www.tungee.com |
| 对应IP | 120.76.128.73 |
| 数字地址 | 2018279497 |
| 物理位置 | 浙江省杭州市 阿里云BGP数据中心 |

二、访问要求： 预计总共需要刷的企业名录240万个，预估需要刷的域名记录是240万\*70%\*2个，按350万预估，需要在7天以内爬完。也就是每天50万的速度。

三、整体目标：

以每天50万的速度，7月24日前爬完所有名单，包括重复爬取（因异常原因第一次未成功爬取的名单）。