

## 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: 企查查网站, <http://www.qichacha.com>;
2. 从数据库读取待查询公司列表, 依次进行查询;
3. 提取搜索结果里的公司的各项信息 (涉及基本信息、法律诉讼、经营状况、对外贸易、企业年报、无形资产) 保存到数据库;
4. 根据“股东”、“分支机构”、“对外投资关系”扩展采集更多的公司信息;
5. 采集频率: 每两个月采集一次;

A	B	C	D	E	F	G	H	I	J	K	L	M	N	
企业ID	企业代码	企业名称	监控类别	所属省份	所属年份	Lng	Lat	url_id	company_name	registration	credit_code	operating_status	organization	company_type
640521	7382831	宁夏天元建材有限公司	废气	宁夏回族自治区	2016	105.69	37.5576	1339b725da68b5c5bc37	宁夏天元建材有限公司	9164052107382831	9164052107382831	开业	7382831-1	一人有限责任公司 (私营)
640381	71509328	青洲英坭水泥有限公司	废气	宁夏回族自治区	2016	105.94	37.8981	1e24b484e6e607c2dc5c	青洲英坭水泥有限公司	6.40000000E+014	916405007E+017	开业	71509328-X	股份有限公司 (非上市)
640402	684211991	宁夏新华实业集团有限公司	废气	宁夏回族自治区	2016	106.27	38.4808	8bacad92917632b27272	宁夏新华实业集团有限公司	--	9.16405007E+017	开业	684211991	一人有限责任公司 (自然人独资)
640521	763202372	宁夏中卫中宁马水泥有限公司	废气	宁夏回族自治区	2016	105.67	37.4318	45c116f97439b52c4f39	宁夏中卫中宁马水泥有限公司	--	9164000076320237	开业	76320237-2	一人有限责任公司 (法人独资)
640202	670412389	国电大武口热电有限公司	废气	宁夏回族自治区	2016	106.39	39.0417	91012020823865555868	国电大武口热电有限公司	--	9164020067041238	开业	67041238-9	有限责任公司 (国有控股)
120225	78316190	天津市晨阳绿源环保科技有限公司	污水处理	天津市	2016	117.44	39.9615	6d5c6d97e32148a3a6754f	天津市晨阳绿源环保科技有限公司	1.20225000E+014	9112022578316190	存续	78316190-X	有限责任公司
640122	22796308	宁夏聚源化工有限公司	废气	宁夏回族自治区	2016	106.28	38.7411	748705480d6f1d0e6704	宁夏聚源化工有限公司	--	6.40122200E+014	开业	22796308-X	一人有限责任公司 (自然人独资)
130208	104364503	唐山冀东水泥股份有限公司 (唐山)	废气	河北省	2016	118.2	39.8678	3529e55748159a3c4b1f	唐山冀东水泥股份有限公司唐山分公司	1.30200300E+014	9113020069755208	存续	69755208-X	台、港、澳投资企业分公司
130223	745410610	宁夏赛马水泥有限公司	废气	宁夏回族自治区	2016	118.67	39.8215	9875ea3663146f1786da	唐山赛马水泥有限公司	--	9.11302237E+017	存续	74541061-0	有限责任公司
640105	546037120	宁夏赛马水泥有限公司	废气	宁夏回族自治区	2016	105.67	37.4318	ac9e2305a3c4d4d1feb0f	宁夏赛马水泥有限公司	--	9164000005460371	开业	05460371-2	一人有限责任公司 (法人独资)
130223	757540439	河北钢铁集团滦县司家营铁矿有限公司	废气	河北省	2016	118.71	39.7727	2429627ccf3aa6e8a1d8f9	河北钢铁集团滦县司家营铁矿有限公司	--	9.11302238E+017	存续	75754043-9	有限责任公司
640121	71507061901	宁夏高元药业有限公司 (建设)	废气	宁夏回族自治区	2016	105.25	38.2805	6aa3e7a090736a890eb2b2	宁夏高元药业有限公司	--	9.16400007E+017	开业	71507061-9	有限责任公司 (自然人独资)
120113	721492376	华鲁恒泰 (天津) 有限公司	重金属	天津市	2016	117.84	39.2276	1e16f06901c1e5902c	华鲁恒泰 (天津) 有限公司	1.20000400E+014	9112011372149237	存续	72149237-6	有限责任公司 (香港法人)
131127	109372847	河北晨化化工有限公司	废气	河北省	2016	116.03	37.6177	9d867ec36530c5cc4219	河北晨化化工有限公司	1.31127700E+014	9.113112717E+017	存续	10937284-7	有限责任公司
131002	601179245	青鸟集团 (唐山) 有限公司	废气	河北省	2016	116.65	39.5115	38c8a8e0b50f5d45d4a9f	青鸟集团 (唐山) 有限公司	--	9.11310005E+017	存续	60117924-5	有限责任公司
130371	745405766	河北天业化工股份有限公司	废气	河北省	2016	115.24	40.4362	bc84e4591e69c83292719a	河北天业化工股份有限公司	--	9113070074540576	存续	74540576-6	股份有限公司 (非上市)
120116	746696884	德玛泰 (天津) 自行车零件有限公司	重金属	天津市	2016	117.41	39.1343	73ae23c4d6d07a9c4971a	德玛泰 (天津) 自行车零件有限公司	1.20000400E+014	9112011674669688	存续	74669688-4	有限责任公司 (外国法人)
130528	741509162	河北健民淀粉糖业有限公司	废水	河北省	2016	114.89	37.6364	804c3c102a754d2815a	河北健民淀粉糖业有限公司	1.30528000E+014	9.11305287E+017	存续	74150916-2	有限责任公司
130523	752420462	建滔 (河北) 焦化有限公司	废水	河北省	2016	114.5	37.1872	257764146c9a06d6a7	建滔 (河北) 焦化有限公司	--	9113050075242046	存续	75242046-2	有限责任公司 (香港法人)
130404	666597261	邯钢集团邯钢钢铁有限公司	废水	河北省	2016	114.47	36.5996	1258526616347265306e	邯钢集团邯钢钢铁有限公司	1.30400000E+014	--	存续	66659726-1	有限责任公司 (法人独资)
130283	755452873	迁安市四通纸业有限公司	废水	河北省	2016	118.73	39.9893	e36ee7a2a6d04949a30	迁安市四通纸业有限公司	1.30283000E+014	9113028375545287	存续	75545287-3	有限责任公司
130207	718339724	唐山市清泉钢铁集团有限责任公司	废气	河北省	2016	118.35	39.5802	52b275e60c14e49e860	唐山市清泉钢铁集团有限责任公司	1.30000000E+014	--	存续	71833972-4	有限责任公司
130204	687038688	唐山冀东水泥股份有限公司	废气	河北省	2016	118.35	39.7585	8c2834793c21a522e0a0	唐山冀东水泥股份有限公司	1.30200000E+014	9.11302007E+017	存续	68703868-8	有限责任公司 (法人独资)
130133	7009545200	河北兴柏药业集团有限公司	废水	河北省	2016	114.89	37.7525	8c3b0f6484bc4bd5c417	河北兴柏药业集团有限公司	1.30133000E+014	--	存续	70095452-0	有限责任公司
130102	6019080221	石药集团中诺药业 (石家庄) 有限公司	废水	河北省	2016	114.53	38.0772	4c489ba2070e28d2e5	石药集团中诺药业 (石家庄) 有限公司	1.30100500E+014	--	存续	60190802-2	台、港、澳投资企业分公司
130105	721677572	中国石化化工股份有限公司沧州分公司	废水	河北省	2016	116.97	38.3372	1e36e2e65d7480630696	中国石化化工股份有限公司沧州分公司	--	9.11309007E+017	存续	72167757-2	股份有限公司分公司 (上市)
130133	7540165300	河北华泰纸业有限公司	废水	河北省	2016	114.76	37.7435	e6d626af7c1b261b6889	河北华泰纸业有限公司	--	9113013375401653	存续	75401653-3	有限责任公司 (法人独资)
120116	600501968	天津士电池有限公司	重金属	天津市	2016	117.7	39.0589	74a902b40d0cb78710e	天津士电池有限公司	1.20000400E+014	9.11201166E+017	存续	60050196-8	有限责任公司 (中外合资)
620422	750938399	甘肃祁连山铝业股份有限公司	废水	甘肃省	2016	105.35	72063	7163309e30487893c	甘肃祁连山铝业股份有限公司	6.20422200E+014	9162042275093839	存续	75093839-9	有限责任公司 (自然人独资)
620502	75655861000	天水华天科技股份有限公司	废水	甘肃省	2016	105.71	34.5747	86cb484e1e3c3292540d	天水华天科技股份有限公司	6.20500000E+014	9162050075655861	存续	75655861-0	股份有限公司 (上市)
640121	715070519	宁夏高元药业 (天津) 有限公司	危险废物	宁夏回族自治区	2016	106.25	38.2805	6aa3e7a74d9036a890eb2b2	宁夏高元药业 (天津) 有限公司	--	9.16400007E+017	开业	71507051-9	有限责任公司 (自然人独资)
120113	600898141	西门子机械传动 (天津) 有限公司	污水处理	天津市	2016	117.07	39.0797	356b9de9804888d159	西门子机械传动 (天津) 有限公司	1.20000400E+014	9112011360089814	存续	60089814-1	有限责任公司 (外国法人)
130733	782560765	崇礼县彭溪排水有限公司	污水处理	河北省	2016	115.27	40.9492	6a2ea5a24bc5c2cc0c	张家口崇礼区彭溪排水有限公司	1.30733000E+014	9113073378256076	存续	78256076-5	其他有限责任公司
620302	224690952	金川集团股份有限公司	废水	甘肃省	2016	102.18	38.5044	6aa3e7a62bdfc0102221e2	金川集团股份有限公司	6.20300000E+014	9162030022469095	存续	22469095-2	股份有限公司 (非上市、国)
620411	750915111	甘肃稀土新材料股份有限公司	危险废物	甘肃省	2016	104.57	36.7025	8e4d4dd8616a3760985e	甘肃稀土新材料股份有限公司	--	9162000075091511	存续	75091511-1	股份有限公司
130729	685712886	万全鑫业水质净化研究中心	污水处理	河北省	2016	114.76	40.7473	0e25c3966e6d9712ba	张家口万全鑫业水质净化研究中心	--	9.11307291E+017	存续	68571288-6	其他所有制
130207	745420296	唐山丰润钢铁 (集团) 有限公司	废气	河北省	2016	118.28	39.4921	14d04a2e0e79d6b2332f	唐山丰润钢铁 (集团) 有限公司	--	9113028274542029	存续	74542029-6	有限责任公司
130683	595430893	河北鑫源环境工程有限公司	污水处理	河北省	2016	114.55	38.0305	89a5c6baa5559917732f	鑫源环境工程有限公司	--	9.11301008E+017	存续	77134423-1	其他有限责任公司
110113	777317074	天津国电津能电力有限公司	废气	天津市	2016	117.36	39.2191	f096a0d0ec7d3e890430f	天津国电津能电力有限公司	1.20110000E+014	9.11201108E+017	存续	77731707-4	有限责任公司 (法人独资)
110113	787750200	北京英商环保有限公司 (顺义)	污水处理	北京市	2016	116.65	40.0355	02357db0c38b329e028d9	北京英商环保有限公司	1.10000010E+014	9111011378775020	开业	78775020-X	其他股份有限公司 (法人独资)
110113	633646910	北京英商环保有限公司	废水	北京市	2016	116.68	40.0965	34d35391827e8d046e90f	北京英商环保有限公司	1.10000010E+014	9111000063364691	开业	63364691-0	其他股份有限公司 (上市)

## 基础信息截图

## 技术难点:

- 企查查系统需要登录才能看到完整信息, 而且登陆过程有阿里云验证码保护, 另外登陆状态每隔2天会被强制失效;
- 企查查封IP很厉害, 容易出现阿里云验证码;

## 实现方案:

- 使用多个账号自动登录采集。登陆过程已实现大多数情况下可以自动完成 (程序能自动拖动网页阿里云验证码), 少数情况下需要人工干预;
- 使用独享稳定高匿HTTP代理绕过封IP限制;

## 项目状态:

本项目于2016年5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](mailto:hello@site-digger.com)

---