

项目总结2023.5

爬取时间为4月24-28日, 分析时间为5月4-5日

涉及单位

目前爬取并分析了阵地类型为官方网站、微信公众号二种类型的数据，其中官网网站58个，微信公众号89个。
其中官方网站爬取了同域名下所有链接地址，微信公众号爬取了历史所有文章。
共爬取网页数量118526页，公众号文章4904篇

分析结果

根据分析要求进行得到分析结果，具体见结果表

存在问题

目前存在部分网站因反爬措施或无法访问或技术原因，未获取到数据，见下表

单位	可能原因
中材高新材料股份有限公司_www.zoomber.com	无法访问
中国国检测试控股集团咸阳有限公司_www.cnmpi.net	爬虫限制
南京国材检测有限公司_www.ctcnj.net	单页面应用
国检测试控股集团安徽元正检测有限公司_www.ctc-yz.com	爬虫限制
中国国检测试控股集团陕西有限公司_www.ceramictest.com	爬虫限制