# 网站建设信息采集系统技术附件

## 概述

[www.egreatwall.com](http://www.egreatwall.com)做为公司业务平台的主要入口，包含有近万家供应商的详细介绍信息和5000多个商品信息，实现与后台业务系统、商机惠等服务平台的对接，随着长城云采2.0系统的开发要进一步做强平台的信息资源，为业务系统开拓市场、挖掘用户创建资源支持平台，为此公司业务会议决定开发“长城快讯”栏目，该栏目主要是从行业的重量级网站中采集信息、抽取信息并按照长城云采2.0门户网站的设计模式发布信息。

## 采集信息源

长城云采门户网站内容维护的业务人员根据业务需求精选了金属材料&冶金、化工原料及制品、泵阀&五金&机电、仪器仪表、机床&机械设备等领域的信息源。金属材料&冶金按照材料划分为钢材、铝材、铜材、型材、型钢等类别，按照信息属性划分为行业动态、企业动态、市场行情和政策法规等不同信息类别；化工原料及制品划分为试剂、助剂、填料、涂料和化学纤维，按照信息性质划分为行业动态、企业动态、行业标准、政策法规、产品动态和技术文章等类别的内容；泵阀&五金&机电划分为泵阀、五金制品、五金工具和机电设备，按照信息性质划分为行业动态、企业动态、行业标准和政策法规；仪器仪表划分为测量仪器、分析仪器、教学仪器、电工仪表、工业仪表和汽车仪表，按照信息性质划分为行业动态、企业动态、行业标准、政策法规、产品动态和技术文章等。信息采集点列表如下所示：

|  |  |
| --- | --- |
| **序号** | **信息采集点名称** |
| 1 | 国家石油和化工网>化工原料及制品>政策法规 |
| 2 | 中国泵阀网>泵阀&五金&机电>行业动态 |
| 3 | 中国泵阀网>泵阀&五金&机电>行业动态 |
| 4 | 全球阀门网>泵阀&五金&机电>行业动态 |
| 5 | 慧聪五金网>泵阀&五金&机电>行业动态 |
| 6 | 全球阀门网>泵阀&五金&机电>企业动态 |
| 7 | 中国泵阀网>泵阀&五金&机电>行业标准 |
| 8 | 全球阀门网>泵阀&五金&机电>政策法规 |
| 9 | 仪器仪表交易网>仪器仪表>行业动态 |
| 10 | 仪器信息网>仪器仪表>企业动态 |
| 11 | 捷配仪器仪表网>仪器仪表>企业动态 |
| 12 | 仪器信息网>仪器仪表>行业标准 |
| 13 | 仪器仪表交易网>仪器仪表>政策法规 |
| 14 | 仪器信息网>仪器仪表>政策法规 |
| 15 | 仪器信息网>仪器仪表>产品动态 |
| 16 | 中国数控机床网>机床&机械设备>行业动态 |
| 17 | 中国机床商务网>机床&机械设备>企业动态 |
| 18 | 中国数控机床网>机床&机械设备>企业动态 |
| 19 | 中国机床商务网>机床&机械设备>产品动态 |
| 20 | 长江有色金属网 >评论分析->金属材料&冶金>行业动态 |
| 21 | 长江有色金属网>行业要闻->金属材料&冶金>行业动态 |
| 22 | 我的钢铁>金属材料&冶金>行业动态 |
| 23 | 长江有色金属网>金属材料&冶金>政策法规 |
| 24 | 我的钢铁>金属材料&冶金>政策法规 |
| 25 | 盖德资讯>化工原料及制品>企业动态 |
| 26 | 国家石油和化工网>化工原料及制品>技术创新 |
| 27 | 中国石油和化工网>化工原料及制品>政策法规 |
| 28 | 国家石油和化工网>化工原料及制品>政策法规 |
| 29 | 工控网>泵阀&五金&机电>行业动态 |
| 30 | 中国泵阀网>泵阀&五金&机电>企业动态 |
| 31 | 中国仪器仪表网>仪器仪表>行业动态 |
| 32 | 仪器信息网>仪器仪表>行业动态 |
| 33 | 仪器仪表网>仪器仪表>行业动态 |
| 34 | 捷配仪器仪表网>仪器仪表>行业动态 |
| 35 | 仪器仪表交易网>仪器仪表>行业标准 |
| 36 | 捷配仪器仪表网>仪器仪表>行业标准 |
| 37 | 仪器仪表网>仪器仪表>产品动态 |
| 38 | 捷配仪器仪表网>仪器仪表>产品动态 |
| 39 | 仪器仪表网>仪器仪表>技术文章 |
| 40 | 捷配仪器仪表网>仪器仪表>技术文章 |
| 41 | 中国机床商务网>机床&机械设备>行业动态 |
| 42 | 中国食品机械网>机床&机械设备>行业动态 |
| 43 | 机经网>机床&机械设备>行业动态 |
| 44 | 机经网>机床&机械设备>企业动态 |
| 45 | 中国食品机械网>机床&机械设备>企业动态 |
| 46 | 中国机械设备网>机床&机械设备>企业动态 |
| 47 | 机经网>机床&机械设备>产品动态 |
| 48 | 中国机床商务网>机床&机械设备>技术文章 |
| 49 | 中国数控机床网>机床&机械设备>技术文章 |
| 50 | 中国机械设备网>机床&机械设备>技术文章 |
| 51 | 机经网>机床&机械设备>技术文章 |
| 52 | 工控网>泵阀&五金&机电>政策法规 |
| 53 | 中钢网>金属材料&冶金>行业动态 |
| 54 | 中钢网>金属材料&冶金>企业动态 |
| 55 | 中钢网>金属材料&冶金>市场行情 |
| 56 | 国家石油和化工网>化工原料及制品>行业动态 |
| 57 | 长江有色金属网>金属材料&冶金>市场行情 |
| 58 | 我的钢铁>金属材料&冶金>市场行情 |
| 59 | 慧聪化工网>化工原料及制品>行业动态 |
| 60 | 盖德资讯>化工原料及制品>行业动态 |
| 61 | 中国石油和化工网>化工原料及制品>行业动态 |
| 62 | 慧聪化工网>化工原料及制品>行情报价 |
| 63 | 2慧聪化工网>化工原料及制品>行情报价 |
| 64 | 3慧聪化工网>化工原料及制品>行情报价 |
| 65 | 4慧聪化工网>化工原料及制品>行情报价 |
| 66 | 5慧聪化工网>化工原料及制品>行情报价 |
| 67 | 6慧聪化工网>化工原料及制品>行情报价 |
| 68 | 盖德资讯>化工原料及制品>行情报价 |
| 69 | 全球石油化工网>化工原料及制品>行情报价 |
| 70 | 慧聪化工网>化工原料及制品>行业统计 |
| 71 | 国家石油和化工网>化工原料及制品>行业统计 |
| 72 | 慧聪化工网>化工原料及制品>企业动态 |
| 73 | 国家石油和化工网>化工原料及制品>企业动态 |
| 74 | 全球石油化工网>化工原料及制品>企业动态 |
| 75 | 盖德资讯>化工原料及制品>技术创新 |
| 76 | 中国石油和化工网>化工原料及制品>技术创新 |
| 77 | 全球石油化工网>化工原料及制品>技术创新 |
| 78 | 慧聪化工网>化工原料及制品>政策法规 |

首先在网站管理系统CMS中设置信息存放栏目，为每个目标栏目定义关键字，CMS系统自动计算被抓取文章与栏目关键字的相关性，相关性最高的栏目获取该篇被抓取的内容，栏目关键字的定义如下表所示：

|  |  |
| --- | --- |
| 51528 | 温度计,压力仪表,燃气表,流量计,水表,热量表,物位,液位,指示仪,记录仪,执行器 |
| 51529 | 速度计,里程表,汽车仪表,油表,转速计,测速仪 |
| 51530 | 测量,测距仪,经纬仪,速测仪,平板仪,水准仪,大地测量,水文,气象,测云仪,温度计 |
| 51531 | 分析仪器,化学分析,分析仪,光度计,照度计,色谱仪,质谱仪,光谱仪,电泳仪,物性分析,气体分析,渗透仪,密度计,浊度计 |
| 51532 | 教学仪表,教学仪器,瓦特表,电压表,电流表,教学 |
| 51533 | 电能表,万用表,示波器,频谱仪,钳形表 |
| 51534 | 温度计,压力仪表,燃气表,流量计,水表,热量表,物位,液位,指示仪,记录仪,执行器 |
| 51535 | 速度计,里程表,汽车仪表,油表,转速计,测速仪 |
| 51536 | 测量,测距仪,经纬仪,速测仪,平板仪,水准仪,大地测量,水文,气象,测云仪,温度计 |
| 51537 | 分析仪器,化学分析,分析仪,光度计,照度计,色谱仪,质谱仪,光谱仪,电泳仪,物性分析,气体分析,渗透仪,密度计,浊度计 |
| 51538 | 教学仪表,教学仪器,瓦特表,电压表,电流表,教学 |
| 51539 | 电能表,万用表,示波器,频谱仪,钳形表 |
| 51540 | 温度计,压力仪表,燃气表,流量计,水表,热量表,物位,液位,指示仪,记录仪,执行器 |
| 51541 | 速度计,里程表,汽车仪表,油表,转速计,测速仪 |
| 51542 | 测量,测距仪,经纬仪,速测仪,平板仪,水准仪,大地测量,水文,气象,测云仪,温度计 |
| 51543 | 分析仪器,化学分析,分析仪,光度计,照度计,色谱仪,质谱仪,光谱仪,电泳仪,物性分析,气体分析,渗透仪,密度计,浊度计 |
| 51544 | 教学仪表,教学仪器,瓦特表,电压表,电流表,教学 |
| 51545 | 电能表,万用表,示波器,频谱仪,钳形表 |
| 51546 | 温度计,压力仪表,燃气表,流量计,水表,热量表,物位,液位,指示仪,记录仪,执行器 |
| 51547 | 速度计,里程表,汽车仪表,油表,转速计,测速仪 |
| 51548 | 测量,测距仪,经纬仪,速测仪,平板仪,水准仪,大地测量,水文,气象,测云仪,温度计 |
| 51549 | 分析仪器,化学分析,分析仪,光度计,照度计,色谱仪,质谱仪,光谱仪,电泳仪,物性分析,气体分析,渗透仪,密度计,浊度计 |
| 51550 | 教学仪表,教学仪器,瓦特表,电压表,电流表,教学 |
| 51551 | 电能表,万用表,示波器,频谱仪,钳形表 |
| 51552 | 温度计,压力仪表,燃气表,流量计,水表,热量表,物位,液位,指示仪,记录仪,执行器 |
| 51553 | 速度计,里程表,汽车仪表,油表,转速计,测速仪 |
| 51557 | 机床,车床,铣床,加工中心,钻床,镗床,磨床,锯床,冲压机,液压机,压力机,折弯机,切削机床,数控机床 |
| 51558 | 挖掘机,推土机,起重机,装载机,压路机,铲运继续,桩工机械,压实机械,工业车辆,摊铺机,叉车,矿山机械,凿岩机,采煤机,破碎机,钻井 |
| 51559 | 制冷,通风,风机,鼓风机,制氢设备,蒸馏,加热设备,干燥设备,浓缩设备,发酵设备,提纯设备,净化设备,离心机,锅炉,电炉,衡器,输送机,升降机,机器人 |
| 51560 | 食品机械,食品设备,食品,制糖,制糖设备,制糖机械,分割机械,分割设备,榨油机,制药机械,制药设备,制药,饲料加工机械,饲料,造纸,造纸设备,造纸机械,制浆,印刷,玻璃,陶瓷,金属冶炼,轧制设备,化工,反应器,反应釜,炼胶机械,注塑机,纺织,织机,印染机械,缝纫机 |
| 51561 | 机床,车床,铣床,加工中心,钻床,镗床,磨床,锯床,冲压机,液压机,压力机,折弯机,切削机床,数控机床 |
| 51562 | 挖掘机,推土机,起重机,装载机,压路机,铲运继续,桩工机械,压实机械,工业车辆,摊铺机,叉车,矿山机械,凿岩机,采煤机,破碎机,钻井 |
| 51563 | 制冷,通风,风机,鼓风机,制氢设备,蒸馏,加热设备,干燥设备,浓缩设备,发酵设备,提纯设备,净化设备,离心机,锅炉,电炉,衡器,输送机,升降机,机器人 |
| 51564 | 食品机械,食品设备,食品,制糖,制糖设备,制糖机械,分割机械,分割设备,榨油机,制药机械,制药设备,制药,饲料加工机械,饲料,造纸,造纸设备,造纸机械,制浆,印刷,玻璃,陶瓷,金属冶炼,轧制设备,化工,反应器,反应釜,炼胶机械,注塑机,纺织,织机,印染机械,缝纫机 |
| 51565 | 机床,车床,铣床,加工中心,钻床,镗床,磨床,锯床,冲压机,液压机,压力机,折弯机,切削机床,数控机床 |
| 51566 | 挖掘机,推土机,起重机,装载机,压路机,铲运继续,桩工机械,压实机械,工业车辆,摊铺机,叉车,矿山机械,凿岩机,采煤机,破碎机,钻井 |
| 51567 | 制冷,通风,风机,鼓风机,制氢设备,蒸馏,加热设备,干燥设备,浓缩设备,发酵设备,提纯设备,净化设备,离心机,锅炉,电炉,衡器,输送机,升降机,机器人 |
| 51568 | 食品机械,食品设备,食品,制糖,制糖设备,制糖机械,分割机械,分割设备,榨油机,制药机械,制药设备,制药,饲料加工机械,饲料,造纸,造纸设备,造纸机械,制浆,印刷,玻璃,陶瓷,金属冶炼,轧制设备,化工,反应器,反应釜,炼胶机械,注塑机,纺织,织机,印染机械,缝纫机 |
| 51620 | 机床,车床,铣床,加工中心,钻床,镗床,磨床,锯床,冲压机,液压机,压力机,折弯机,切削机床,数控机床 |
| 51621 | 挖掘机,推土机,起重机,装载机,压路机,铲运继续,桩工机械,压实机械,工业车辆,摊铺机,叉车,矿山机械,凿岩机,采煤机,破碎机,钻井 |
| 51622 | 制冷,通风,风机,鼓风机,制氢设备,蒸馏,加热设备,干燥设备,浓缩设备,发酵设备,提纯设备,净化设备,离心机,锅炉,电炉,衡器,输送机,升降机,机器人 |
| 51623 | 食品机械,食品设备,食品,制糖,制糖设备,制糖机械,分割机械,分割设备,榨油机,制药机械,制药设备,制药,饲料加工机械,饲料,造纸,造纸设备,造纸机械,制浆,印刷,玻璃,陶瓷,金属冶炼,轧制设备,化工,反应器,反应釜,炼胶机械,注塑机,纺织,织机,印染机械,缝纫机 |
| 51426 | 钢材,粗钢,钢铁,黑色系,特钢,不锈钢 |
| 51427 | 铝,铝材,铝型材,电解铝,氧化铝,铝土矿,铝棒,铝板,铝合金 |
| 51428 | 铜,铜材,电解铜,铜精矿,废铜,铜管,铜板带,铜棒,铜合金 |
| 51429 | 铝型材,型材 |
| 51430 | 型钢,热轧,冷轧,中厚板,H型,钢带,钢管 |
| 51431 | 焦炭,废钢,钢坯,铁矿石,炉料 |
| 51432 | 钢材,粗钢,钢铁,黑色系,特钢,不锈钢 |
| 51433 | 铝,铝材,铝型材,电解铝,氧化铝,铝土矿,铝棒,铝板,铝合金 |
| 51434 | 铜,铜材,电解铜,铜精矿,废铜,铜管,铜板带,铜棒,铜合金 |
| 51435 | 铝型材,型材 |
| 51436 | 型钢,热轧,冷轧,中厚板,H型,钢带,钢管 |
| 51437 | 焦炭,废钢,钢坯,铁矿石,炉料 |
| 51438 | 钢材,粗钢,钢铁,黑色系,特钢,不锈钢 |
| 51439 | 铝,铝材,铝型材,电解铝,氧化铝,铝土矿,铝棒,铝板,铝合金 |
| 51440 | 铜,铜材,电解铜,铜精矿,废铜,铜管,铜板带,铜棒,铜合金 |
| 51441 | 铝型材,型材 |
| 51442 | 型钢,热轧,冷轧,中厚板,H型,钢带,钢管 |
| 51443 | 焦炭,废钢,钢坯,铁矿石,炉料 |
| 51444 | 钢材,粗钢,钢铁,黑色系,特钢,不锈钢 |
| 51445 | 铝,铝材,铝型材,电解铝,氧化铝,铝土矿,铝棒,铝板,铝合金 |
| 51446 | 铜,铜材,电解铜,铜精矿,废铜,铜管,铜板带,铜棒,铜合金 |
| 51447 | 铝型材,型材 |
| 51448 | 型钢,热轧,冷轧,中厚板,H型,钢带,钢管 |
| 51449 | 焦炭,废钢,钢坯,铁矿石,炉料 |
| 51456 | 试剂,显色剂,染色剂,指示剂,缓冲剂,干燥剂,高纯物 |
| 51457 | 助剂,处理剂,添加剂,加工剂,改性剂,稳定剂,润滑剂,催化剂,引发剂,乳化剂,分散剂,阻聚剂,终止剂,氧化剂,调节剂,扩链剂,中和剂,脱水剂,固化剂,交联剂,引发剂,稀释剂,阻燃剂,增强剂,发泡剂,促进剂,软化剂,除味剂 |
| 51458 | 涂料，家具漆，水性漆，漆，腻子 |
| 51459 | 填料，空心球，滤料,排水帽,瓷球,鲍尔环,填充剂,增量剂,密封胶,拉西环 |
| 51460 | 化纤,纤,纶,棉纶,涤纶,粘纤,铜氨纤,腈纶,氨纶,维纶,丙纶 |
| 51461 | 合成材料,塑料,橡胶,树脂,乙烯,丙烯,丁烯,甲醇,乙醇,丁醇,苯,苯酚 |
| 51474 | 试剂,显色剂,染色剂,指示剂,缓冲剂,干燥剂,高纯物 |
| 51475 | 助剂,处理剂,添加剂,加工剂,改性剂,稳定剂,润滑剂,催化剂,引发剂,乳化剂,分散剂,阻聚剂,终止剂,氧化剂,调节剂,扩链剂,中和剂,脱水剂,固化剂,交联剂,引发剂,稀释剂,阻燃剂,增强剂,发泡剂,促进剂,软化剂,除味剂 |
| 51476 | 涂料，家具漆，水性漆，漆，腻子 |
| 51477 | 填料，空心球，滤料,排水帽,瓷球,鲍尔环,填充剂,增量剂,密封胶,拉西环 |
| 51478 | 化纤,纤,纶,棉纶,涤纶,粘纤,铜氨纤,腈纶,氨纶,维纶,丙纶 |
| 51479 | 合成材料,塑料,橡胶,树脂,乙烯,丙烯,丁烯,甲醇,乙醇,丁醇,苯,苯酚 |
| 51462 | 试剂,显色剂,染色剂,指示剂,缓冲剂,干燥剂,高纯物 |
| 51463 | 助剂,处理剂,添加剂,加工剂,改性剂,稳定剂,润滑剂,催化剂,引发剂,乳化剂,分散剂,阻聚剂,终止剂,氧化剂,调节剂,扩链剂,中和剂,脱水剂,固化剂,交联剂,引发剂,稀释剂,阻燃剂,增强剂,发泡剂,促进剂,软化剂,除味剂 |
| 51464 | 涂料，家具漆，水性漆，漆，腻子 |
| 51465 | 填料，空心球，滤料,排水帽,瓷球,鲍尔环,填充剂,增量剂,密封胶,拉西环 |
| 51466 | 化纤,纤,纶,棉纶,涤纶,粘纤,铜氨纤,腈纶,氨纶,维纶,丙纶 |
| 51467 | 合成材料,塑料,橡胶,树脂,乙烯,丙烯,丁烯,甲醇,乙醇,丁醇,苯,苯酚 |
| 51486 | 试剂,显色剂,染色剂,指示剂,缓冲剂,干燥剂,高纯物 |
| 51487 | 助剂,处理剂,添加剂,加工剂,改性剂,稳定剂,润滑剂,催化剂,引发剂,乳化剂,分散剂,阻聚剂,终止剂,氧化剂,调节剂,扩链剂,中和剂,脱水剂,固化剂,交联剂,引发剂,稀释剂,阻燃剂,增强剂,发泡剂,促进剂,软化剂,除味剂 |
| 51488 | 涂料，家具漆，水性漆，漆，腻子 |
| 51489 | 填料，空心球，滤料,排水帽,瓷球,鲍尔环,填充剂,增量剂,密封胶,拉西环 |
| 51490 | 化纤,纤,纶,棉纶,涤纶,粘纤,铜氨纤,腈纶,氨纶,维纶,丙纶 |
| 51491 | 合成材料,塑料,橡胶,树脂,乙烯,丙烯,丁烯,甲醇,乙醇,丁醇,苯,苯酚 |
| 51468 | 试剂,显色剂,染色剂,指示剂,缓冲剂,干燥剂,高纯物 |
| 51469 | 助剂,处理剂,添加剂,加工剂,改性剂,稳定剂,润滑剂,催化剂,引发剂,乳化剂,分散剂,阻聚剂,终止剂,氧化剂,调节剂,扩链剂,中和剂,脱水剂,固化剂,交联剂,引发剂,稀释剂,阻燃剂,增强剂,发泡剂,促进剂,软化剂,除味剂 |
| 51470 | 涂料，家具漆，水性漆，漆，腻子 |
| 51471 | 填料，空心球，滤料,排水帽,瓷球,鲍尔环,填充剂,增量剂,密封胶,拉西环 |
| 51472 | 化纤,纤,纶,棉纶,涤纶,粘纤,铜氨纤,腈纶,氨纶,维纶,丙纶 |
| 51473 | 合成材料,塑料,橡胶,树脂,乙烯,丙烯,丁烯,甲醇,乙醇,丁醇,苯,苯酚 |
| 51480 | 试剂,显色剂,染色剂,指示剂,缓冲剂,干燥剂,高纯物 |
| 51481 | 助剂,处理剂,添加剂,加工剂,改性剂,稳定剂,润滑剂,催化剂,引发剂,乳化剂,分散剂,阻聚剂,终止剂,氧化剂,调节剂,扩链剂,中和剂,脱水剂,固化剂,交联剂,引发剂,稀释剂,阻燃剂,增强剂,发泡剂,促进剂,软化剂,除味剂 |
| 51482 | 涂料，家具漆，水性漆，漆，腻子 |
| 51483 | 填料，空心球，滤料,排水帽,瓷球,鲍尔环,填充剂,增量剂,密封胶,拉西环 |
| 51484 | 化纤,纤,纶,棉纶,涤纶,粘纤,铜氨纤,腈纶,氨纶,维纶,丙纶 |
| 51485 | 合成材料,塑料,橡胶,树脂,乙烯,丙烯,丁烯,甲醇,乙醇,丁醇,苯,苯酚 |
| 51496 | 泵,阀,水泵,油泵,化工泵,离心泵,真空泵,潜水泵,污水泵,球阀,蝶阀,截止阀,止回阀,调节阀,角阀,泵阀,真空阀,调压阀 |
| 51497 | 日用五金,建筑五金,锁,合页,铰链,索具,滑轮,水暖,脚手架,五金配件,螺丝,螺母,螺栓,齿轮,轴承,弹簧,滑道,炊具,锅,丝网,固件,磨料,搪瓷 |
| 51498 | 锯,钳子,扳手,螺丝刀,五金工具,工具,钻,黄油枪,锤,尺,剪,刀 |
| 51499 | 电机,发电机,电动机,工控,电工,电气,减速机,变压器,电控,电抗器,开关,连接器,继电器,控制台 |
| 51500 | 泵,阀,水泵,油泵,化工泵,离心泵,真空泵,潜水泵,污水泵,球阀,蝶阀,截止阀,止回阀,调节阀,角阀,泵阀,真空阀,调压阀 |
| 51501 | 日用五金,建筑五金,锁,合页,铰链,索具,滑轮,水暖,脚手架,五金配件,螺丝,螺母,螺栓,齿轮,轴承,弹簧,滑道,炊具,锅,丝网,固件,磨料,搪瓷 |
| 51502 | 锯,钳子,扳手,螺丝刀,五金工具,工具,钻,黄油枪,锤,尺,剪,刀 |
| 51503 | 电机,发电机,电动机,工控,电工,电气,减速机,变压器,电控,电抗器,开关,连接器,继电器,控制台 |
| 51504 | 泵,阀,水泵,油泵,化工泵,离心泵,真空泵,潜水泵,污水泵,球阀,蝶阀,截止阀,止回阀,调节阀,角阀,泵阀,真空阀,调压阀 |
| 51505 | 日用五金,建筑五金,锁,合页,铰链,索具,滑轮,水暖,脚手架,五金配件,螺丝,螺母,螺栓,齿轮,轴承,弹簧,滑道,炊具,锅,丝网,固件,磨料,搪瓷 |
| 51506 | 锯,钳子,扳手,螺丝刀,五金工具,工具,钻,黄油枪,锤,尺,剪,刀 |
| 51507 | 电机,发电机,电动机,工控,电工,电气,减速机,变压器,电控,电抗器,开关,连接器,继电器,控制台 |
| 51508 | 泵,阀,水泵,油泵,化工泵,离心泵,真空泵,潜水泵,污水泵,球阀,蝶阀,截止阀,止回阀,调节阀,角阀,泵阀,真空阀,调压阀 |
| 51509 | 日用五金,建筑五金,锁,合页,铰链,索具,滑轮,水暖,脚手架,五金配件,螺丝,螺母,螺栓,齿轮,轴承,弹簧,滑道,炊具,锅,丝网,固件,磨料,搪瓷 |
| 51510 | 锯,钳子,扳手,螺丝刀,五金工具,工具,钻,黄油枪,锤,尺,剪,刀 |
| 51511 | 电机,发电机,电动机,工控,电工,电气,减速机,变压器,电控,电抗器,开关,连接器,继电器,控制台 |
| 51518 | 测量,测距仪,经纬仪,速测仪,平板仪,水准仪,大地测量,水文,气象,测云仪,温度计 |
| 51519 | 分析仪器,化学分析,分析仪,光度计,照度计,色谱仪,质谱仪,光谱仪,电泳仪,物性分析,气体分析,渗透仪,密度计,浊度计 |
| 51520 | 教学仪表,教学仪器,瓦特表,电压表,电流表,教学 |
| 51521 | 电能表,万用表,示波器,频谱仪,钳形表 |
| 51522 | 温度计,压力仪表,燃气表,流量计,水表,热量表,物位,液位,指示仪,记录仪,执行器 |
| 51523 | 速度计,里程表,汽车仪表,油表,转速计,测速仪 |
| 51524 | 测量,测距仪,经纬仪,速测仪,平板仪,水准仪,大地测量,水文,气象,测云仪,温度计 |
| 51525 | 分析仪器,化学分析,分析仪,光度计,照度计,色谱仪,质谱仪,光谱仪,电泳仪,物性分析,气体分析,渗透仪,密度计,浊度计 |
| 51526 | 教学仪表,教学仪器,瓦特表,电压表,电流表,教学 |
| 51527 | 电能表,万用表,示波器,频谱仪,钳形表 |

## 信息采集系统

实现从互联网上采集信息的技术手段是采用“网络机器人技术”，该技术实现了网页的自动浏览和信息抓取，并可以对抓取的信息实现“有用信息”的抽取，将抽取到的信息直接保存到网站信息发布系统中，通过信息发布系统进行再次发布，在企业内部门户网站上和企业外部宣传网站上展现从互联网上获取的信息。

众所周知，互联网的基本协议是HTTP协议，在网站上展示信息的主要语言是HTML语言，HTTP协议定义了服务器与浏览器之间进行信息交换的标准，HTML语言定义网站信息的组织形式。一个网站可以看成是一个按照树型结构组织的信息集合，信息之间的连接靠的是URL连接来传递。HTTP协议是一种无状态协议，浏览器向WEB服务器发出信息请求后，WEB服务器向浏览器返回它所要获取的信息，一个会话结束，浏览器和服务器之间并没有状态信息被保存，HTTP协议定义了一组标准的会话命令，最常使用的命令是GET命令和POST命令，其中网页上的URL连接是采用GET命令，网页上的FORM表既可以采用的GET命令，也可以采用POST命令与WEB服务器进行会话，HTTP协议的另一个命令是HEAD命令，该命令不获取网页的信息体，只是获取网页的头信息，一般在自动浏览网页的网络机器人中经常使用，主要是用于检查网页的状态。

将“采集信息源”列表中定义的采集点配置到信息采集系统中，采集源配置列表如下所示：



每个信息采集点需要配置URL匹配正则表达式和网页信息抽取特征码，根据URL正则表达式确定采集那些地址的信息，通过配置信息抽取特征码确定从具体的网页内容页抽取那些可用的信息。

配置好采集信息源后需要对每个信息源进行信息采集测试，全部信息源采集测试完成后可以开启信息采集，系统采用不间断循环的方式进行信息采集，系统同时启动设置的线程数采用多线程方式进行信息采集，可以同时对多个网站的内容进行信息采集。



## 采集信息发布模板开发

金属材料&冶金按照材料划分为钢材、铝材、铜材、型材、型钢等类别，按照信息属性划分为行业动态、企业动态、市场行情和政策法规等不同信息类别；化工原料及制品划分为试剂、助剂、填料、涂料和化学纤维，按照信息性质划分为行业动态、企业动态、行业标准、政策法规、产品动态和技术文章等类别的内容；泵阀&五金&机电划分为泵阀、五金制品、五金工具和机电设备，按照信息性质划分为行业动态、企业动态、行业标准和政策法规；仪器仪表划分为测量仪器、分析仪器、教学仪器、电工仪表、工业仪表和汽车仪表，按照信息性质划分为行业动态、企业动态、行业标准、政策法规、产品动态和技术文章等

|  |  |  |  |
| --- | --- | --- | --- |
| **一级栏目** | **二级栏目** | **二级栏目** |  |
| 金属材料&冶金 | 钢材 | 行业动态 |  |
| 铝材 | 企业动态 |  |
| 铜材 | 市场行情 |  |
| 型材 | 政策法规 |  |
| 型钢 |  |  |
| 化工原料及制品 | 试剂 | 行业动态 |  |
| 助剂 | 企业动态 |  |
| 填料 | 行业标准 |  |
| 涂料 | 政策法规 |  |
| 化学纤维 | 产品动态 |  |
|  | 技术文章 |  |
| 泵阀&五金&机电 | 泵阀 | 行业动态 |  |
| 五金制品 | 企业动态 |  |
| 五金工具 | 行业标准 |  |
| 机电设备 | 政策法规 |  |
| 仪器仪表 | 测量仪器 | 行业动态 |  |
| 分析仪器 | 企业动态 |  |
| 教学仪器 | 行业标准 |  |
| 电工仪表 | 政策法规 |  |
| 工业仪表 | 产品动态 |  |
| 汽车仪表 | 技术文章 |  |
| 机床&机械设备 |  |  |  |

## 长城信息采集系统的运行环境

采集软件访问地址：<http://47.93.212.40/spider/>

登录用户名：admin/!1

运行服务器外部地址47.93.212.40（内部172.17.139.40），Java1.8版本以上，应用服务器为resin-4.0以上，使用ORACLE数据库，resin的配置如下所示：

<!--- JSSE default properties-->

<system-property

jdk.tls.ephemeralDHKeySize="2048"

jdk.tls.rejectClientInitiatedRenegotiation="true"

sun.security.ssl.allowUnsafeRenegotiation="false"

sun.security.ssl.allowLegacyHelloMessages="false"/>

<database>

<jndi-name>jdbc/spider</jndi-name>

<driver type="oracle.jdbc.driver.OracleDriver">

<url>jdbc:oracle:thin:@localhost:1521:orcl11g</url>

<user>egwdbadmin</user>

<password>qazwsxokm</password>

</driver>

<prepared-statement-cache-size>8</prepared-statement-cache-size>

<max-connections>20</max-connections>

<max-idle-time>30s</max-idle-time>

</database>

<cluster id="app">

<host id="" root-directory=".">

<web-app id="/spider" root-directory="软件安装目录">

<class-loader>

<compiling-loader path="WEB-INF/classes"/>

<library-loader path="WEB-INF/lib"/>

</class-loader>

</web-app>

</host>

</cluster>