# 工作周报

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **汇报日期** | | 2017/08/018 | **汇报人** | 李静 | |
| **汇报对象** | | 安徽征信股份有限公司（梁龙） | | | |
|  | | | | | |
| **本周工作内容** | | | | | |
| **序号** | **工作内容** | | | | **完成情况** |
| 1 | 国家企业信用信息公示系统(全国)企业信息的爬取方案 | | | | >100% |
| 2 | 编写企业页面信息的抓取脚本，抓取信息的保存结构设计 | | | | >60% |
| 3 | 优化代码 | | | | >30% |
|  | | | | | |
| **本周遇到的难题** | | | | | |
| 1.全国版的信息里夹带扰乱字符随机生成，每次打开网页都不同，去除扰乱字符 | | | | | |
| 2.字段为图片的，图片是一串URI，它就是小图片本身，无法下载识别，分析网页得到相同名称字段的URI是相同的，从网页抓取URI，得到URI和字段的映射，后面考虑用判断得到的URI的方式，确定图片字段名称 | | | | | |
|  | | | | | |
| **下周工作计划** | | | | | |
| **序号** | **计划内容** | | | | |
| 1 | 抓取信息的保存结构设计 | | | | |
| 2 | 完善页面脚本的编写 | | | | |